Hemagglutinin-based antigen designs expressed in the methylotrophic yeast *Pichia pastoris*

by

José Arquímides Echánove Juan

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Chemical Sciences

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne

Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	Hemagglutinin-based antigen designs expressed in the methylotrophic yeast Pichia pastoris
Name of Candidate Nom du candidat	Echánove Juan, José Arquímides
Degree Diplôme	Master of Science
Department/Program Département/Programme	Chemical Sciences
Date of Defence Date de la soutenance	December 19, 2016

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Robert Lafrenie
(Co-Supervisor/Co-directeur(trice) de thèse)

Dr. Francisco Diaz-Mitoma
(Co-Supervisor/Co-directeur(trice) de thèse)

Dr. Eric Gauthier
(Committee member/Membre du comité)

Dr. Yan Zhou
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. Shelley Watson
Madame Shelley Watson
Acting Dean, Faculty of Graduate Studies
Doyenne intérimaire, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Abstract

The Influenza virus is a significant global health problem whose worldwide circulation constantly triggers new pandemic strains often linked to genetic re-assortment and new mutations. Virus variability threatens and challenges the health of millions of citizens. Vaccination against influenza prior to infection offers the standard alternative to antiviral treatment in infected patients. Vaccine production must be available in sufficient volume for human and animal vaccination on a large scale. In this study I designed a novel recombinant influenza protein composed of three chimeric Hemagglutinin (HA) influenza receptor-binding domains (RBD) expressed in tandem. This construct was expressed as a single molecule in *Pichia pastoris* and characterized by gene sequencing, polyacrylamide gel electrophoresis, immuno-detection, and carbohydrate composition. The goal of this project is to develop a reagent for use in immunological response assays in a mouse model that will provide information regarding the potential use of these multisubunit gene-expressing constructs as vaccine candidates.

Keywords

Pichia pastoris, Influenza virus, Hemagglutinin, Antigen, Protein Expression, Gene Design, Receptor Binding Domain, Vaccination.

Acknowledgments

I really want to thanks all the Academic personal at Laurentian University, including my professors and fellows, for a great experience in Sudbury, Ontario. Special thanks to Dr. Francisco Diaz Mitoma for inviting me to come to Canada to pursuit graduate studies at Laurentian Universirty, using the laboratories of Health Sciences North Research Institute (formerly AMRIC), and for his academic support. To Dr Robert Lafrenie for helping me during the last year of my Master study, for his free spirit to make things easier, just like life must be in a scientific enviroment.

To Dr Hoang-Thanh Le, Dr. Reza Nokhbeh, Dr. Justo Perez, Dr. Gustavo Ybazeta, Dr. Manuel Aguilar, Dr. Jordan Lewicky and Chef Brad Davidson (special merit), for their unconditional support and advisory during the realization of this work, thank you for all your time.

To my family who really understands the reason of my passion for science. Monica, Vania, Valentina. I love you all. This is a beautiful journey, for all of us.

Table of Contents

List of Tables

List of Figures

List of Appendices

1. INTRODUCTION

Influenza virus infection is a global epidemic that each year affects more than 100 million people worldwide, causing a serious health situation in 3-5 million persons, and an estimated 250,000-500,000 deaths (*1*). Global concern about the development of periodic new pathogenic strains, which have the capacity to create new pandemics, requires that influenza outbreaks are constantly being monitored by the scientific community in centres for surveillance, hospitals and at points of care. The “Spanish Flu’ global pandemic started prior to 1920 and killed more than 50 million people over a 50 year timespan. The novel “swine influenza” rapidly circulated across the World in 2009 causing 15,000 death (*1*). Attempts to prevent further influenza pandemic require an improvement in the vaccination methods and in the technology necessary to control virus-associated illness. Data from Public Health Agency of Canada for the year 2014-2015 reveals the seasonal persistence of influenza A and B types, with most cases belonging to the influenza strain AH3N1 subtype, followed by Influenza B type infections, and then by the AH1N1 pandemic strain (2,3) (Figure 1). Seasonal variation, as reported by the World Health Organization (Figure 2 and 3), differs significantly depending on the type of virus circulating in the different transmission zones (*4*). Paradoxically, the efficiency by which people can move or connect across the planet makes us more prone to the efficient dispersal of these pathogens.

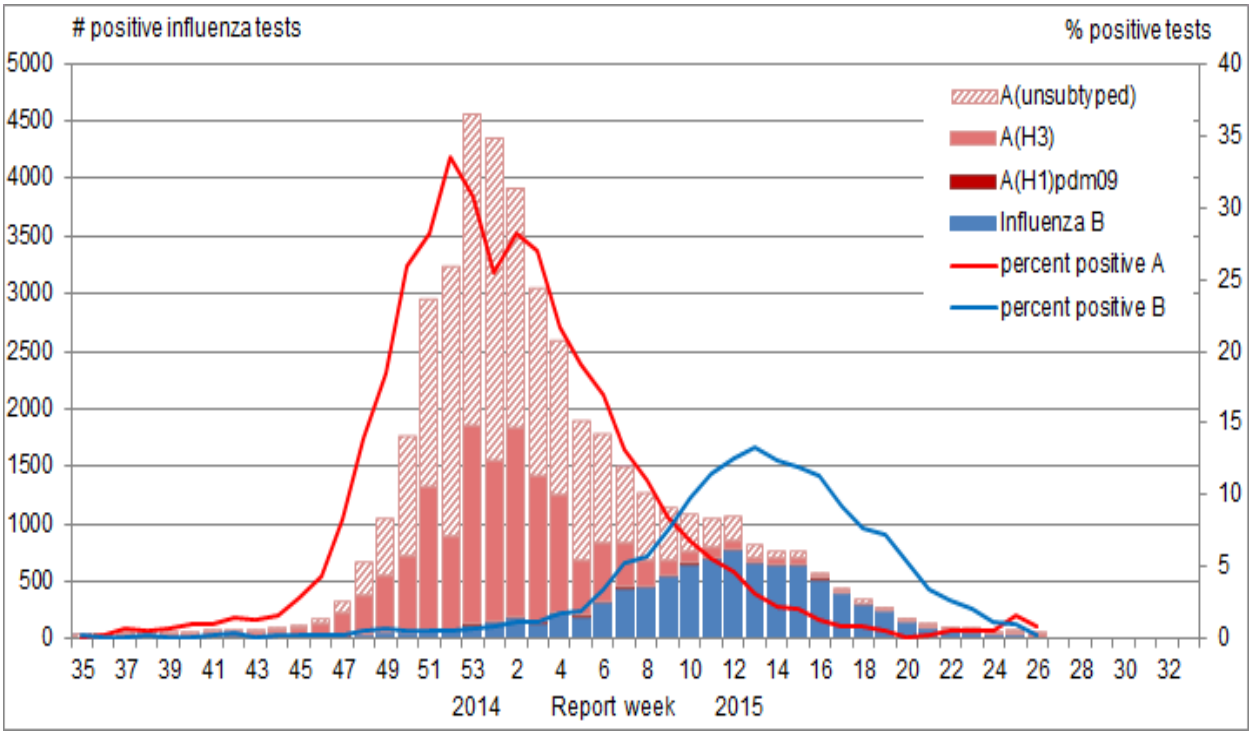


Figure 1. Number the number of positive influenza tests and percentage of tests positive, by type, subtype, and report week for, Canada, 2014-15 (Public Health Agency of Canada). Source:©All Rights Reserved. *Weekly Influenza Reports.* Public Health Agency of Canada, 2016. Adapted and reproduced with permission from the Ministery of Health, 2016.

1.1. Influenza disease

1.1.1 Virus composition and structure (Influenza A Type, Influenz B type)

Influenza viruses are negative single-stranded RNA-based viruses [(–)ssRNA (family Orthomyxoviridae)] that are classified in five genera, *Influenzavirus A, Influenzavirus B*, *Influenzavirus C*, *Thogovirus,* and *Isavirus*, that differ in composition and have differences in protein sequence. Influenza A type is subtyped based on the antigenic differences in the Hemagglutinin (HA) and Neuroaminidase (NA) proteins. Sixteen HA (HA1-16) subtypes and nine NA (N1-N2) subtypes have been described to date. In contrast, no similar description of the antigenic subtypes has been found for the Influenza B and C types (*5*).

The influenza virus is composed of a lipid membrane envelope derived from the host cell. Influenza A and B types share some characteristics, for example, both of their genomes are composed of 8 RNA segments that each codes for at least one protein. In the case of the influenza A virus, the eighth ssRNA segment encodes for 12 known proteins, divided into nine structural proteins and three non-structural proteins as summarized in Table 1 (1,*6,7*).

The influenza A virus exhibits a remarkable pleomorphic structure with shapes varying from spherical structures in the range of 80-120 nm in diameter (*5*), to filamentous forms isolated from fresh clinical samples. This phenotypic pleomorphism is commonly found in the novel AH1N1 2009 pandemic strain (*8,9,10*) and is promoted mainly by changes in the amino acid sequence of the M1 inner envelope protein that contributes to viral pathogenesis (*11*).

Table 1. Genome composition of the influenza virus: the segmented genes involved, their translated protein products, and their functions in the life cycle of the Influenza A respiratory virus (1,*6,7*).

Structural proteins

Segment 1	Polymerase Basic Protein 2 (PB2)	Temperature-dependent replication competence; mRNA synthesis by binding host mRNA caps. vRNP complex.
Segment 2	Polymerase Basic Protein 1 (PB1)	Component of the RNA polymerase involved in RNA replication and transcription. vRNP complex.
Segment 3	Protein Associated (PA)	Component of the RNA polymerase involved in RNA replication and transcription. Make the vRNP complex.
Segment 4	Hemagglutinin (HA)	Binding and fusion with the host cell (sialic acid linkage specificity); antigenicity and tropism.
Segment 5	Nucleoprotein (NP)	Enclose the viral genome in an helical organized fashion (Chenavas et al 2013).
Segment 6	Neuroaminidase (NA)	Receptor-destroying enzyme. Confers resistense to antivirals (oseltamavir). Promotes efficient release of virion from infected cells.
Segment 7	Matrix protein 1 (M1)	Component of the internal viral envelope.
Segment 7	Matrix protein 2 (M2)	Ion channel. Component of the internal viral envelope along with M1. Confers resistance to antivirals (adamantanes).
Segment 8	Nuclear Export Protein 2 NS2 (NEP)	Localized in virions facilitating nuclear export of viral RNP complex. Facilitate activity-encreased.

Non-structural proteins

Segment 2	PB1-F2	Induction of apoptosis; promotion of secondary bacterial infection.
Segment 2	N40	Function not fully understood.
Segment 8	Non-structural protein 1(NS1)	Host antiviral response antagonist.

1.1.2. Hemagglutinin is the major antigenic molecule of influenza

Hemagglutinin is a Type I integral membrane glycoprotein that is part of the viral lipid bilayer surface architecture forming spikes-like structures that binds to host cell receptors by interacting with sialic acid moieties (12,13). The hemagglutinin protein is encoded by segment four of the viral genome and has a protein size of 60 KDa prior to post-translational modification. In the Influenza A virus, biologically active Hemagglutinin is a homotrimer with each monomer comprised of two domains: the globular domain and the stem domain. Structurally, the globular domain folds into a jelly roll-like motif with eight anti-parallel beta sheets forming a pocket known as the receptor-binding site that interacts with carbohydrates containing alpha-2-6-linked or alpha-2-3-linked sialic acid moieties (7,13,14). It also possesses a vestigial esterase domain that is shared with the stem domain and which mainly forms a long alpha coiled-coil structure which plays an important role in membrane fusion during virion entry into the plasma membrane of the host cell (12). The structure of the stem domain consist of a cleavage signal, a hydrophobic peptide domain that is involved in membrane fusion, and a C-terminal transmembrane domain (*12*–*14*). A disulfide bond is formed between cysteine residues in the globular and stem domains which stabilizes the molecular conformation after the proteolytic cleavage that occurs during hemagglutinin fusion peptide exposure and full activity (*12*).

The hemagglutinin molecule is the major antigen involved in generating most of the neutralizing antibodies produced in response to infection, with the majority of the antigenic sites located within the globular domain (*12*). Therefore, the globular domain of the hemagglutinin molecule is the target for immune protection generated during

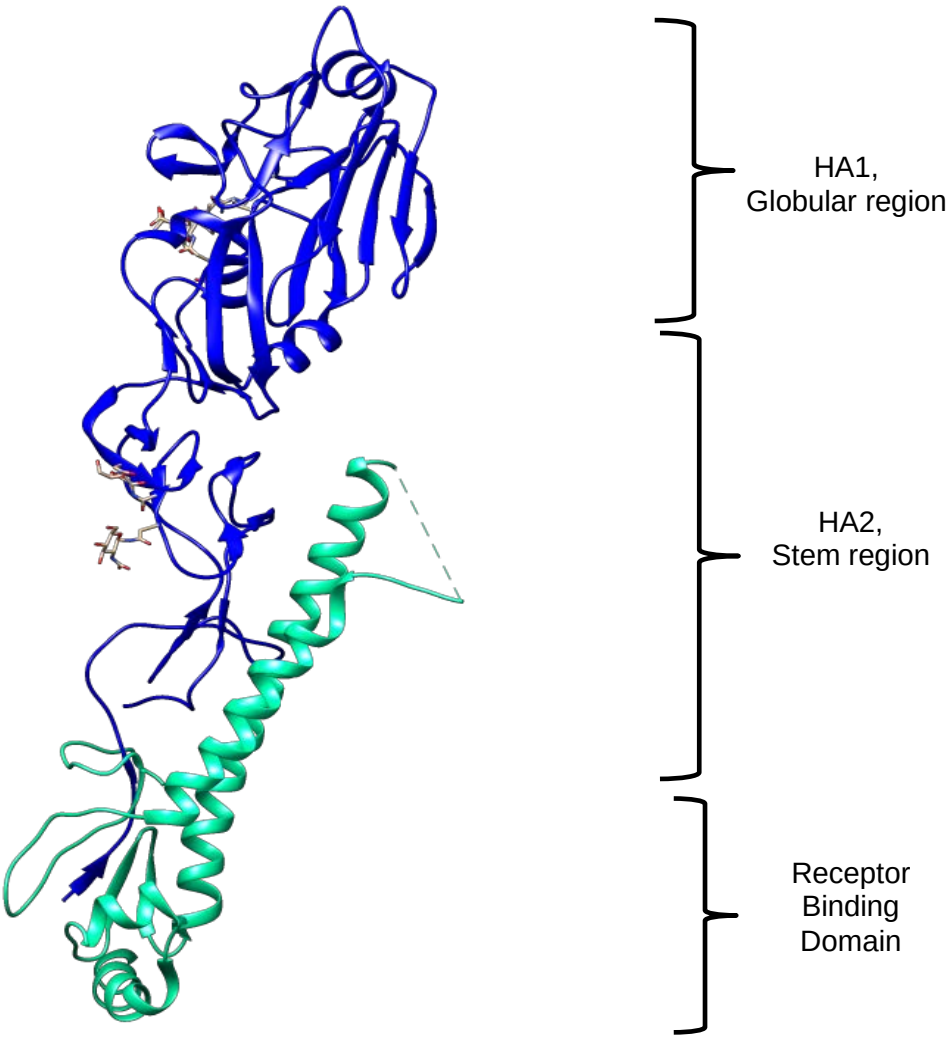


Figure 2. The basic structure of Influenza Hemagglutinin (AH1N1 2009 pandemic virus). After proteolytic cleavage, the HA0 molecule is divided in the HA1 (in blue) subunit or globular region which harbors the lectin- or RBD-domain, and the HA2 (in green) subunit, or stem region. The structure is based on the H1 subtype of the virus A/H1N1 2009 pdm (*16*) (RCSB Protein Data Bank (PDB) ID: 4DEA) using PyMOL-Chimera UCSF for visualization .

infection in humans, other mammals and avian species. Consequently, hemagglutinin has been been selected as the main target in the development of vaccines, and inhibitor molecules (15), and constitutes the target for antigen design and protein expression in the present study.

1.1.3. Mechanisms of influenza infection

The early events of influenza infection, involve the binding of the hemagglutinin globular domain to cellular receptors via the alpha-2-3-linked and/or alpha-2-6-linked sialic acid carbohydrates which promotes entry into the cell membrane by clathrin-dependent (*17*) and independent endocytosis (*18*). Cellular proteases are responsible for hemagglutinin cleavage which exposes the fusion peptide and promotes fusion between the viral envelope and the endosomal membrane. Maturation of the endosome further activates the virus. The acidification of the endosomal environment has two main consequences: i) it activates Matrix Protein 2 ion channels that acidify the inside of the virion particle, un-coating the RNP complex from the viral genome; and, ii) it promotes the cytoplasmic release of the virion genome originated by the pH-dependent fusion of hemagglutinin to the endosomal membrane (*1*). The genetic information is then translocated to the nucleus of the cell where the RNA-dependent RNA polymerase replicates a series of RNA molecules to create the negative-sense RNA molecules that are necessary templates for packaging co-assembly with the viral proteins. Complementary, positive sense RNA segments, are exported to the cytoplasm to be translated to create the viral protein and to be used in the synthesis of the negative sense small viral RNA (svRNAs) involved in regulating the switch between the transcription and replication processes of the viral messenger RNA in the cytoplasm

(*1,19*). Once translation of the mRNAs has started, the NP and M1 proteins return to the nucleus and interact with the negative sense vRNA (*1,20*). The synthetized proteins produced in the nucleus, can be post-translationally processed in the Golgi apparatus, and directed towards the plasma membrane (HA, NA, M2) where M1 helps to assemble the virus particle (*1,21*), Release of the membrane-encapsulated virion particle from the plasma membrane occurs via facilitation of the Neuraminidase enzymatic activity which cleaves the sialic acid moieties (*22*).

1.1.4. Virus tropism

The ability of the influenza virus to infect more than one species is known as tropism. For example, the H1N1 and H3N2 virus are exclusive in humans, birds and pigs (*23*), because their hemagglutinin surface proteins have the ability to recognize different sialic acid molecules on the cells from many species. The viral receptors are tissue-specific and are expressed among different types of hosts. For example, Swine virus recognizes a receptor containing alpha-2-3-linked and alpha-2-6-linked sialic acids, both presented on swine trachea, which promotes the binding of the HA molecule to the swine and human upper respiratory tract-rich alpha-2-6 linked sialic acid receptors and lower respiratory-rich alpha-2-3-linked sialic acid receptors (*1,24*). The Avian virus infects birds preferentially using alpha-2-3-linked sialic acid receptors present in their guts and respiratory systems but can also bind to swine alpha 2-3-linked sialic acid, and human 2-3-linked sialic acid receptors (*23,25*). Receptor switching can occur by mutagenesis, as in the case of the avian H3 subtype, where a single change in two residues confers an affinity for human 2-6-linked sialic acid receptors (*26*). Thus, a major shift in the Hemagglutinin and Neuraminidase segments among distantly related

viruses, can generate significant opportunities for the newly arranged virus combinations to infect humans and spread faster on the surface of host cells. For example, structural studies on the Receptor Binding domains RBDs of HA have demonstrated that amino acid substitution within this region is responsible for receptor specificity and tissue tropism. For the H2 and H3 subtypes, the amino acid mutations Q226L and G228S in the RBD confer binding preference to switch from avian to human receptors (*27*). In another study, using ferret differentiated primary cells, the human influenza virus was able to infect ciliated epithelial cells whereas the Highly Pathogenic Avian Influenza Virus, H5N1, was able to replicate efficiently in non-conciliated epithelial cells where α2,3-linked sialic acid receptor are highly abundant (*28*).

1.1.5. Virus reassortment and antigenic drift as a major source of antigenic variation

Virus re-assortment occurs when two or more Influenza viruses of the same type are able to co-infect a single host cell and exchange their segmented genomes (*5)*. This major genome exchange contributes to the generation of pandemic virus reassortants like the recent Influenza A H1N1 2009 pandemic outbreak (*29*). H1N1 2009 was comprised of a triple re-assortant between Avian Swine and Human genomes, that coexisted in the naturally occurring swine influenza A virus strains, which increased the efficiency of virus replication in human lungs cells and enhanced aeroseol transmissibility (*30*).

Antigenic drift in influenza viruses is the result of genetic modification in the influenza virus genome, that result from point mutations that accumulate mainly due to error – prone replication of the viral genome viral RNA-dependent polymerase. These

mutations can become fixed in the viral genome and can then be selected for their effect on virulence by selection primarily of the Hemagglutinin and Neuraminidase genes which promote viral spread. The result of this selection can be creation of virulence- and immunologically-distinct viruses that can efficiently spread in the population.

1.2. FLU vaccine technologis: Overview

Influenza vaccines represent the best current method to prevent pathogenesis resulting from influenza virus infection. Conventional and approved methods for vaccine production are based on the production of attenuated virus particles following virus inoculation, culture and harvest using chicken eggs (*31*). Each year, new vaccine formulations are engineered and millions of single dose vaccines are created for human and animal immunization programs (*32*). The yearly vaccine formulations are based on the predictions of multiple surveillance programs which predict which virus serotypes will be in circulation and cause disease (*33*). Scientists around the world work hard and invest huge amounts of time and resources to create vaccines that correspond to the most likely combinations of virus strains and re-assorted viral forms that are present in the world each year. The drawbacks of this approach are that different viruses serotypes can emerge or unperceived pandemic strains can circumvent the implemented detection strategies (*31*) or the ability control them (*34*). When unpredicted influenza subtypes emerge they can create a significant health problem which cannot be met by the vaccines created by the conventional egg-dependent virus production

methodologies which require a significant amount of time. Although this method represents a stable and reliable vaccine technology, it takes a significant amount of time to produce a new vaccine formulation. Current alternative technologies may be developed to better compete in such hypothetical but plausible pandemics events (*31,34–36*).

1.2.1. Current vaccine technologies

1.2.1.1. Propagation of Influenza Virus in Embryonated Chicken Eggs

The production of new virus for influenza vaccines candidates is based on the propagation of multiple virus antigens in embryonated chicken eggs and requires the combination or reassortment of the identified genes from the most likely influenza virus strains. To create the vaccine, the genes of the first, target flu strain, are expressed in an egg with the genes of a second flu strain that grows well in eggs but is harmless to humans (i.e. a virus that does not replicate at human body temperature). Initially, both flu strains are injected into a fertilized chicken egg and both genomes are replicated. During their replication, the influenza genes segments mix and reassemble to create new viral offspring combinations. Since each genome contains 8 different gene segments, this reassortment makes it possible to have as many as 256 different genes combinations in the viral offspring. As a consequence, a labor intensive screening process is required to identify the desired features from both strains, for example, the expression of surface antigens corresponding to the new pandemic Hemagglutinin and Neuroaminidase proteins integrated into the genome of the egg-adapted and replication-defective (to humans) second virus (*37*).

The conventional and approved method of virus antigen production for vaccine composition may be some complicated when the target strains infect the eggs with lower efficiency making it difficult to create the desired levels of final antigen product (*31*). There is also a high risk of cross contamination of the vaccine with material isolated from the egg during vaccine manufacture (*38*). In addition, there are more complex facilities required for vaccine manufacture using this method. These various complications are driving the search for a new method to produce appropriate influenza antigens that can be adopted for the production of vaccine.

1.2.1.2. Influenza DNA-based vaccines

The formulation of DNA vaccines is based on using purified DNA that encodes the appropriate viral proteins. The viral DNA is present in plasmid vectors, which can be electroporated into host cells to exploit the host cellular machinery to translate the specific antigens and express them in the transformed animal tissues (*39,40*). Viral DNA-based vaccines are able to stimulate humoral, cell-mediated responses. For example, vaccination with a plasmid that codes for a viral antigen in combination with a plasmid encoding a neutralizing monoclonal antibody for the same antigen, is able to induce a synergistic effect and a broader immune response, compared to separate administration, representing a more robust and flexible vaccine platform (*41–44*).

1.2.1.3. Influenza Synthetic Peptide Vaccines

Influenza vaccines have also been creating using synthetic peptides. This approach allows the use of chemically synthetized peptides that correspond to portions of the Influenza proteins that are well known to elicit immune protection via innate or

humoral responses. This methodology has made it possible to create many formulations targeting different epitopes of the virus, with peptide cocktails based on conserved or variable regions and with the ability to generate immune cross-protection among various types or subtypes of virus in the same formulation (*45–47*). However, since smaller peptide antigens do not stimulate strong immune responses these formulations also have to include adjuvants to stimulate the immune response which have to be evaluated to produce effective vaccine candidates (*48*) .

1.2.1.4. Universal Vaccines

The creation of Universal vaccines requires the use of antigens that correspond to highly conserved regions of the virus and which code for proteins critical for the function of the pathogen's infection process. For example, the main targets of universal influenza vaccines include epitopes that are critical for the conformational changes necessary for the persistence and success of infection (*49*). In general, peptide sequences that have been conserved during the evolution of the pathogen and that play a critical function in the mechanism of infection are the most successful candidates. Evolution of the virus sequence is also important in terms of immune system recognition and neutralization since protection of critical sites of the molecules is required to ensure survival of the virus. Examples where viral evolution has resulting in burying those sequences within a larger macromolecule are very well known. In HIV, conserved sequences in the gp120 molecule which correspond to the CD4-binding site (CD4bs) and coreceptor-binding site (CoRbs) are only exposed after conformational changes occur in gp120 during infection. This means that the immune system is unable to normally make antibodies which block gp120-mediated infection (*50*). The presence of

antibodies can be used for the identification of such epitopes, since epitope mapping with neutralizing polyclonal or monoclonal antibodies can identify crucial regions that might make functionally important vaccine targets. Therefore, the composition of a universal vaccine can be based on using short or larger peptides with linear conformational structure, peptides that correspond to conserved or consensus critical regions, or a set of engineered peptides that present a stabilized conserved antigen that might be functionally important (*49*).

1.2.1.5. Recombinant subunit vaccines. using yeast Pichia pastoris as a platform for vaccine production for novel influenza antigens.

The methylotrophic yeast *Pichia pastoris* has been of great interest and widely used for the production of vaccines since the technology arose in the middle of the 1980's (*51*). Its versatility relies on many factors:

a) *Pichias pastoris* is an eukaryotic microorganism that is very easy to culture and which requires no-complex media formulations for its growth and maintenance. The *Pinchia pastoris* optimal growth condition are very well established with an optimal temperature of 28-30^{o}C that permits the expression of temperature-stable proteins and has the advantage that its growth kinetics may not be significantly altered if the optimum temperature for recombinant protein expression has to be lowered to the range of 20-24^{o}C. *Pichia pastoris* can also generate a large biomass even in the absence of protein induction, which makes it suitable for fermentative bioprocessing.
b) *The expression of the heterologous genes can be placed under the control of very strong DNA promoters*. The Alcohol Oxidase 1 (***AOX1***) promoter has a

DNA size of 960 base pairs, and drives the expression of polypeptides using methanol as the sole inducer carbon source (52). The Glyceraldehyde-3-Phosphate promoter, or its alternate mutated form, is 477 DNA base pairs in length and can be used to enhance protein expression (53,54). Furthermore, expressing a combination of two different recombinant protein-promoter constructs inserted in the same genome of a *Pichia pastoris* clone is used as a biotechnological strategy to increase protein expression synergistically compared to when they are used separately (personal experience).

c) *The genome of Pichia pastoris can be modified with standard molecular biology techniques*. The exogenous genes cloned in a *Pinchia pastoris* expression plasmid, can be presented as a linear or circular vector inserted into the yeast genome using chemical, electrical, physical, or the combination of all the aforementioned transfection methods. For example, the pPicZalpha plasmids can be successfully integrated into the *Pinchia pastoris* genome by homologous recombination at the terminal sites of the digested AOX1 promoter, thus facilitating the integration of multiple DNA cassettes during a single transformation procedure. The selection process employs an antibiotic resistance marker, like the Zeocin resistant *ble* gene or the blasticidin S-resistance (*bsr*) gene. As a rule, the amount of soluble protein expressedby a single transformed Pichia clone can be increased by using higher levels of selecting antibiotics to isolate antibiotic hyper-resistance (up to 2000 µg/mL) clones, which is an indirect indicator of clones where multiple cassettes containing the gene of interest have been integrated.

d) Recombinant protein *expression and posttranslational modification* in *Pichia pastoris is robust where other expression systems have failed*. Expression of therapeutic proteins, antigens, monoclonal antibodies, and peptides in *Pichia pastoris* are abundant in the literature ranging from milligram to grams per liter (55,56,57). The possibility of expressing the recombinant protein at significant levels with large industrial volumes in bioreactors, must be considered an important step when the production of antigens for human or animal vaccination programs is being proposed. The influenza hemagglutinin subunits have been successfully expressed in *Pichia pastoris* cells at high levels and in a conformation capable of eliciting production of neutralising antibodies in animals (56). On the other hand *Pichia pastoris* possesses an eukaryotic posttranslational modification system that is useful when the proteins require a certain glycosylation pattern for expression of the full activity of human-intended therapeutic proteins (58) . The normal forms of N-linked or O-linked glycosilation motif patterns are easily recognized in this yeast system. *Pichia pastoris* can introduce hyper-glycosylated carbohydrate side chains (59,60) that may affect protein functionality, folding, or protein-protein interactions (61). In addition, *Pichia pastoris* strains can be bioengineered to generate humanized N-glycosylation patterns (59),(62). Substitution of residues within the glycosylation motifs by DNA mutagenesis has also been shown to be a method to generate protein heterogeneity without affecting the biological function that is commonly associated with hyper-glycosylation (61,63). Expression of recombinant proteins in prokaryotic systems results in un-glycosylated proteins that can have an

impact on antibody responses in animal models, which can affect success (63,64) or cause low efficacy (65) of antibody binding in response to high virus challenges. This deficiency in the effectiveness of prokaryote-derived proteins is often associated with incorrect conformational folding and lack of post-translational modification that may affect the formation of the neutralising epitopes (64), (65). The intention of the present study is focused more on the advantages of the eukaryotic yeast expression system for a robust yield of novel influenza antigens to assist in the production of influenza globular head domains for heterosubtypic immunity or to be integrated in approaches to induce stem-reactive antibodies (*49,63*)

1.3. Protein engineering and antigen design

A perfect vaccine should comprise many characteristics in its design:

a) **Gene-based protein vaccines should stimulate the immune system.** The larger the protein molecule, the better the immune response (66). In this way the use of adjuvants is abate: not meaning that they are no longer needed but instead, that their addition increases the chance that less antigen in the final formulation can still stimulate a strong immune response, thus resulting in more doses of vaccines being produced (67).

b) **The expressed protein molecules should resemble the normal 3-dimensional structure, folding and post-translational modification of the antigen protein**. There is some evidence that post-translational modifications,

like glycosylation modification, may negatively affect the immune response although this concept has been challenged. For instance, the enzymatic de-glycosylation of Hamagglutinin proteins expressed in cell cultures, are able to induce a better neutralizing serum protection than untreated proteins in vaccinated organisms (65). Therefore, changing the expression platform from a eukaryotic to a prokaryotic system results in an antigen that lacks glycan moieties that can elicit high neutralizing antibodies titers (68). In addition, bioengineering of cysteine residues in the hemagluttinin stem-based domain to create peptides that showed conformational constraints to mimic the proper antigenic exposition and promote stabilization by disulphide bridging enhanced its effectiveness (69,70).

c) **The antigen must be produced at high levels with the protein expression system used and must be easy to purify**. Some antigens express better by changing the expression system used, changing from prokaryotic to eukaryotic or *vice versa* may overcome such situations. Strategies like adding a soluble domain and a protease cleavage site next to the C- or N- terminal portion of the gene of interest can facilitate the recovery of the protein from the soluble phase followed by its release from the unwanted domain. Even though protein precipitation like the situation which occurs in inclusion bodies in prokaryotic systems permits the recovery of large amounts of antigen, the refolding process is not always successful at restoring immunogenicity. It is also likely that the current use of bioinformatics tools for codon optimization of the gene, based on

the relative abundance of transfer RNA and secondary mRNA structure, can be an important optimizing parameter for heterologous expression.

1.4. HYPOTHESIS AND AIMS

The main objective of this study is to to obtain expression of chimeric Receptor Binding domains in *Pichia pastoris* without resulting in insoluble proteins as previously reported in prokaryotic systems.

To generate a platform for protein expression in *Pichia pastoris* that can be used to produce high levels of recombinant influenza virus proteins that are composed of at least of three Receptor Binding Domains.

The receptor binding domains represents antigenic determinants in different Influenza types and subtypes, in this regards, a single expressed antigen can be used as a vaccine against hetero-subtyped potential candidate viruses.

Evaluate the recovery of native designed antigens and tagged designed antigens that help to simplify the purification steps for future scale-up conditions.

Characterize the antigens expressed in Pichia pastoris by its biochemical methods (immunodetection, glycan-content, N-teminal sequencing, DNA sequencing, Mass-Spec analsysis) that helps us to asses and confirm the identity of the antigens expressed.

Based on expression and characterization results, apply the use of bioinformatics that help us to improve the antigens design.

2.0 METHODS

2.1. Influenza RBD expression in *Pichia pastoris*

2.1.1 Gene design for avian and human RBD trimers

2.1.1.1. Gene sequences and synthesis.

Influenza full HA sequences of avian origin were obtained from the biological manufacturer company, Investigacion Aplicada S.A. (IASA). These sequences represent influenza virus isolates from Mexican chicken farms, which have been identified according to the strain of origin. The sequences of the DNA were aligned using the Basic Local Aligment Search Tool (BLAST) from the National Center for Biotechnology Information (NCBI) to show areas of homology. Protein sequences corresponding to the HA molecules of human origin where retrieved from the NCBI database based on the 2013-2014 influenza vaccine composition as approved for use by the World Health Organization (71).

The antigen design was based on previous studies, which have used the Receptor Binding Domain from Influenza AH1N1 California 2009 pandemic, named $RBS_{(28\text{-}265),}$ as the basis for a candidate vaccine (64,68). This approach uses these lectin globular domains from different virus-specific HA1 subunits as the antigenic molecules for the creation of a potential multivalent vaccine. These antigens lack the fusion peptide sequence and other amino acid residues related to the stalk, and HA2 subdomain. The selected sequences were modelled individually using the Phyre2 website server to calculate the predicted globular structure (72).

The gene construct designed for this vaccine was composed of three RBD's expressed in tandem (the RBD trimer) with two flexible linker sequences that code for (serine-glycine-glycine-glycine-glycine)2 (10 a.a. length) to facilitate an intermolecular spacing between each of the monomers. The Avian RBD trimer gene version V3.4 is composed of the three avian RBD: (5'-3' DNA sense): an Avian H5-like Hemagglutinin domain, followed by another Avian H5-like hemagglutinin domain, and an Avian H7-like hemmagglutinin virus domain.

Based on the Avian RBD trimer V3.4, three additional versions were designed: V3.5, V3.6 and V3.7. These constructs included the addition of an HIV-Tat basic domain cell-penetrating peptide, an immune-stimulant 5-mer peptide, or a combination of both (immune-stimulant 5-mer peptide and HIV-Tat basic domain cell-penetrating peptide), respectively, engineered on the 3' end of the coding region.

The Human RBD trimer version V3.3 is composed of the following RBDs (listed in the 5'-3' DNA sense): the Human A type, H1 hemmagglutinin domain from the AH1N1/California/09/2009 pandemic strain; the Human A type, H3 hemmaglutinin domain from AH3N2 A/Texas/50/2012 virus; and, the Human B type, B hemagglutinin domain from the B/Brisbane/33/2008 virus.

The sequences of the artificial gene RBD trimer versions were code-optimized and synthetized for the *Pichia pastoris* expression system using the services of Genescript Inc (Piscataway, NJ). The gene construct versions were originally designed to contain an XhoI (5' *CTCGAG* 3') restriction site at the 5' end of the gene and an XbaI (5' *TCTAGA* 3') restriction site followed by a Stop codon (TAA) at the 3' end of the gene

in all the versions. The two restriction sites allow for unidirectional cloning in the *Pichia pastoris* extracellular expression plasmid pPicZalphaA.

2.1.1.2. Custom Cloning

The different versions of the genes (V3.3, V3.4, V3.5, V3.6, and V3.7) were synthethized containing a stop codon at the C-terminus of the artificial RBD trimers and were custom cloned in the extracellular expression plasmid pPicZalphaA (**Figure 3**) using the two restriction sites XhoI and XbaI. The first restriction site (XhoI) is in frame with the sequences encoding the endopeptidase cleaveage site, Kex2, thus facilitating the complete removal of the alpha factor peptide and leaving the native N-terminal portion of the RB trimer without any additional modification (**Figure 4**).

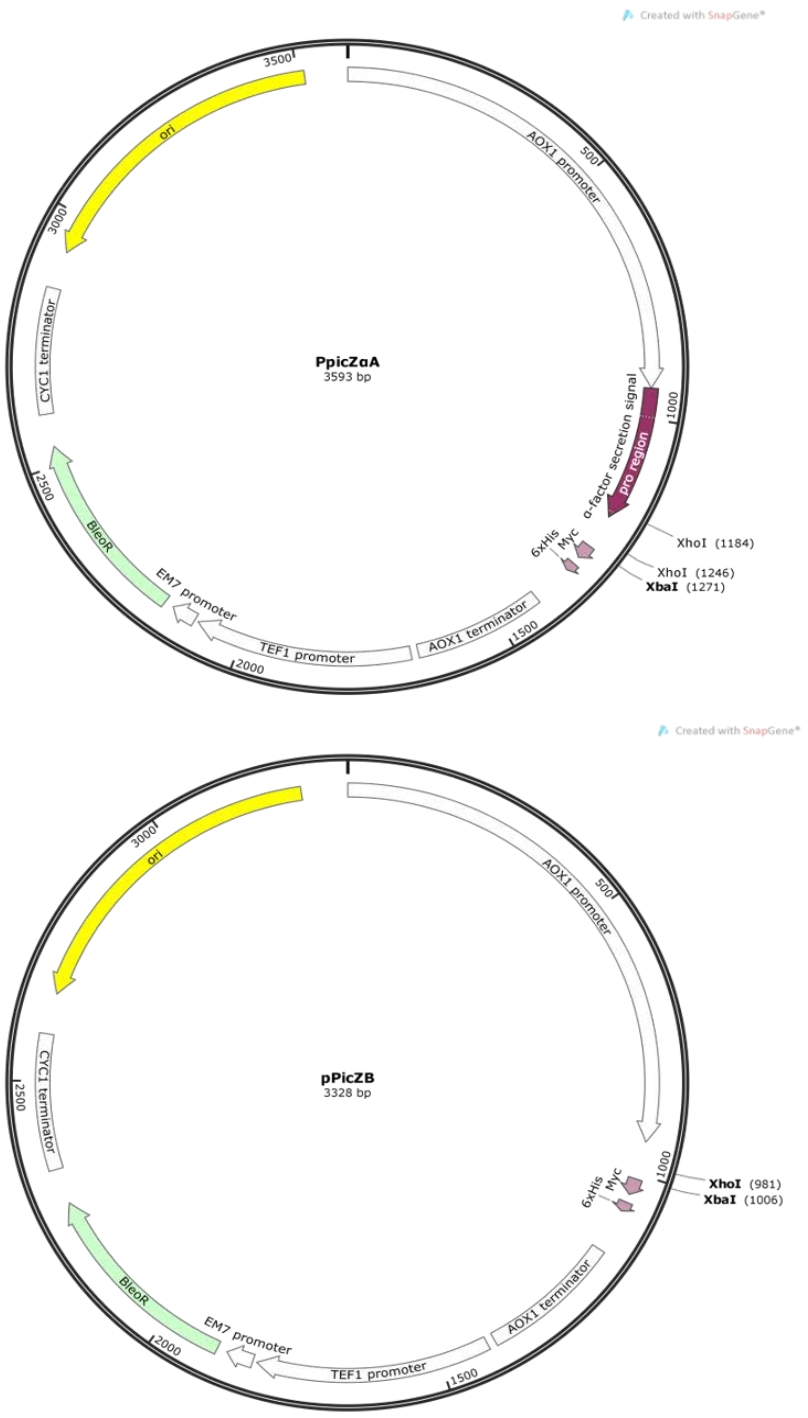


Figure 3 . The extracellular pPicZ-Alpha-A and Intracellular pPicZ-B vectors used for the expression of Avian and Human HA-RBD synthetic and codon-optimized versions in the *Pichia pastoris* expression system. Two restriction sites, XhoI and XbaI were used to clone the HA RBD multimeric genes. The extracellular vector possesses the alpha factor secretion signal for protein export.

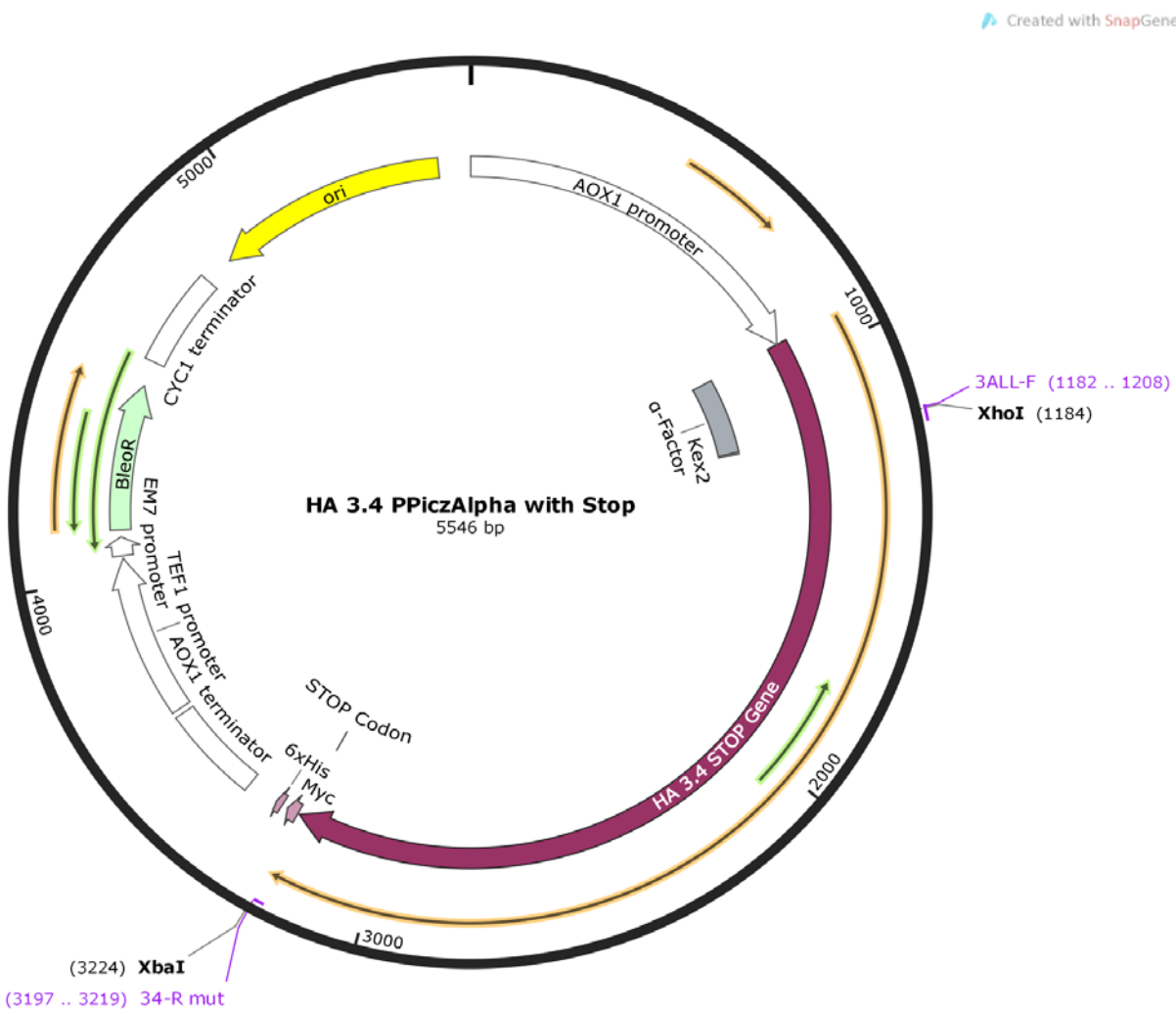


Figure 4 . General custom cloning strategy for expression of extracellular Avian and Human HA-RBD multimeric genes in the plasmid pPicZalpha-A. In this particular case, Avian HA-RBD V3.4 was synthetized (GeneScript, Inc.) containing a premature stop codon after the C-terminal sequence of the multimeric RBD gene. Alpha factor sequences facilitate export to the secretory pathway and removal by the Kex2 endopeptidase. The intention of this first design strategy was to produce and to purify recombinant RBDs in their native sequence state.

2.2. Modifications to gene design

2.2.1. Removal of premature stop codon in the Avian and Human HA-RBD versions to restore 6-Myc and 6-His tag

As part of a second strategy, the modification of the codon optimized synthetic Avian and Human HA-RBD versions was performed. A PCR amplification, using specific primers, removed the stop codon located after the C-terminal portion of the HA-RBDs. This allowed the creation of an open reading frame that included the c-Myc and 6-His tag sequences for purification and detection purposes that were present in the original plasmid design (**Figure 5**). At the same time, the avian and human HA-RBD versions were inserted into the pPicZ-B vector to explore the effects of intracellular expression of the various RBD proteins in *Pichia pastoris*.

The PpicZ-B plasmid lacks the in-frame reading of the alpha-factor signal peptide which is required to secrete the recombinant protein. The cloned Avian and Human RBD genes (V3.3, V3.4, V3.6, V3.7) in the pPicZ-alphaA plasmids were used as the template in a PCR reaction using an antisense primer that altered the premature stop codon to facilitate the in-frame translation of the c-myc and 6-His tags. Briefly, a specific primer corresponding to the 5' DNA region was used to add an XhoI restriction site sequence, a yeast consensus Kozak's sequence and a methionine start codon sequence to the 5' sequence of each of the avian RBD gene constructs in order to facilitate the correct ligation, translation initiation and intracellular expression of the RBD trimer proteins. The new constructs were named: **HA-RBD mutV3.3**, **HA-RBD mutV3.4**, **HA-RBD mutV3.6**, and **HA-RBD mutV3.7**.

2.2.2. PCR mutagenesis

To clone into the pPicZalpha-A plasmid, the Avian RBD gene constructs were isolated using PCR performed. The PCR used the forward primer 3ALL-F and the reverse primer 3.4RMUT for amplification of version V3.4; the forward primer 3ALL-F and the reverse primer 3.6-Rmut were used to amplify version 3.6; the forward primer 3ALL-F and the reverse primer 3.7Rmut were used for version V3.7. Additionally, RBD's from human origin V3.3 were amplified using the forward primer 33-FWD and the reverse primer 33-R and subsequently cloned in the extracellular expressing vector pPicZalpha-A.

For cloning of the Avian RBD into the intracellular expression vector pPicZ-B, the forward primer INTRA-F and the reverse primer 3.4Rmut were used for PCR amplification of version V3.4. The forward primer INTRA-F and the reverse primer 3.6-Rmut were used for the amplification of version V3.6 The forward primer INTRA-F and the reverse primer 3.7Rmut were used to amplify version V3.7. The human RBD version mutV3.3 was PCR amplified using the primers F-INTRA33 and 33-R and 200 ng of Pichia genomic DNA from clone D7 which had been transformed with the inserted pPicZalpha-A HA Human RBD V3.3 construct (premature Stop codon-No 6His tagged). Primers are shown in **Table 2.**

The PCR reactions consisted a total volume of 50 μl containing 5 ng of either pPicZalphaA plasmid vector containing one of the unmodified versions of the Avian RBD contructs V.3.4, V3.6, V3.7; forward primer (0.5 μM final concentration), reverse primer (0.5 μM final concentration), 1 μl of High Fidelity Phusion Polymerase (1U total)

(ThermoFisher), 1 μl of dNTPs (10 mM each), 10 μl of Buffer HF (1x final concentration) and H_2O. Cycling parameters consisted of an initial denaturation step (98^oC for 30 sec); followed by 30 amplification cycles (each cycle consisted of a denaturation step at 98^oC for 10 sec, a primer annealing-amplification step at 72^oC for 1 min 10 sec; and a final extension at 72^oC for 10 min); and a final incubation step at 4^oC applied to all samples until they were removed.

All reverse primers were designed to avoid amplification of the stop codon located next the 3' site of the Avian RBD template version but kept the XbaI restriction site for re-cloning into the pPicZalpha-A and pPicZB expression vectors (**Figure 2**). Two forward primers were designed to amplify the sense DNA strand: for all avian RBD versions, 3ALL-F 5' GAGGAGGAGCTCTCGAGAAGAGAAGTGTTAAGGGAG 3' was used; and for the human RBD version the forward primer 33-Fwd 5' GAGGAGGAGCTCTCGAGAAAAGAGGTGTCGCTCCATTGC 3' and the reverse primer 33-R 5' GAGGAGGAGCTTCTAGACCTTTGATGACTTTACTTCTACC 3' was used in the PCR amplification for cloning in the extracellular vector pPicZalpha-A.

2.2.3. Avian and Human HA-RBD re-cloning in pPicZalpha and pPicZ-B

The PCR of each version of the RBD amplified a ds DNA fragment of approximately 2.1 Kb. These bands were excised and recovered after electrophroresis on 1.2 % agarose gels in TAE 1X, pH 8.0 (40mM Tris, 20mM acetic acid, and 1mM EDTA) using the Qiaquick Gel extraction kit (Qiagen Inc) following the manufacturer recommendation. The isolated fragments were quantitated by spectroscopy at 260/280 nm and 500 ng of purified PCR products were double-digested at 37^oC for 4 h with both

the XhoI (30 Units total) and XbaI (30 Units total) restriction enzymes in 1X Cutsmart Buffer in 50 μl total reaction volume. After incubation, the samples were heat inactivated at 70^{o}C for 20 min, and subjected to electrophoresis on a 1.2% Agarose gel in 1X TAE buffer along with the DNA standard 1kb DNA ladder (Invitrogen). Samples were visualized under a UV trans-illuminator and the DNA fragments that correspond to the RBD's (2.1 Kb long) were recovered from the agarose gel and stored in 1.5 ml microcentrifuge tubes. The gel slices were extracted using the gel recovery kit (Qiagen) following the manufactured instructions and the recovered DNA was quantified using a Nanodrop 2000c spectrophotometer evaluating the purity at 260/280 nm.

The DNA vectors pPicZalpha and pPicZ-B were double-digested in 50 μl total volume reaction using 500 ng of each plasmid and the restriction enzymes XhoI (10 units total) and Xba (10 units total) in 1X Cutsmart Buffer and incubated to 37^{o}C for 4 h. After incubation, the samples were subjected to electrophoresis on a 1.2% agarose gel in 1X TAE buffer along with the DNA standard 1kb DNA ladder (Invitrogen). Samples were visualized under a UV transilluminator and the DNA fragments that correspond to the pPiczAlpha-A (3.6 Kbases long) or pPicZ-B (3.1 kbases long) plasmids were collected separately by excising the gel slices, which were stored in 1.5 mL microcentrifuge tubes. The gel slices were extracted using the gel recovery kit (Qiagen) following the manufactured instructions and the recovered DNA was quantitated using a Nanodrop 2000c spectrophotometer at 260/280 nm.

For DNA ligation, a 3:1 insert to vector molar ratio was used. The reaction was performed using 87 ng of DNA corresponding to each RBD gene version (insert), and 50 ng each of pPicZalpha-A or pPicZ-B digested DNA (the inserts were cloned

separately into both plasmid vectors) with 1 μl of T4 ligase enzyme (0.5 Units, Fermentas, Fisher Scientific) and 1X T4 ligase buffer (Fermentas,). The mixture was combined gently using a pipet tip and incubated at 23^{o}C for 1 h followed by incubation at 4^{o}C for 4 h. The ligation reaction was stopped by incubation at 70^{o}C for 20 min to inactivate the enzyme and then the ligation reactions were stored at -20^{o}C or used immediately for DNA transformation.

The plasmids created by ligating the RBD genes into the pPicZalpha-A or pPicZ-B vector backbones were transformed into *E.coli* using a heat shock protocol. 50 μl of *E. coli* TOP10 chemical competent cells (in 60 mM calcium chloride, 15% (w/v) glycerol, 10 mM PIPES, pH 7.0, solution) were incubated with 2 μl of ligation reaction in an ice-bath for 30 min, then the cells were heat shocked at 42^{o}C for 45 s, and immediately returned to the ice bath for 2 min. The cells were recovered by incubation in pre-warmed (37^{o}C) LB-Low-Salt, pH 7.5 medium (1% Tryptone, 0.5% NaCl, 0.5% yeast extract) for 1 h. 100 μl of the culture was then spread on LB-agar plates containing the Zeocin (25 μg s/ml) selective agent and incubated at 37^{o}C for 12 h to allow clonal growth. Five to ten positive *E. coli* clones were picked individually and grown in 10 mL of LB-low-salt pH 7.5 liquid media containing Zeocin (25 μg/mL) for plasmid selection. The cultures were grown for 12-14 h at 37^{o}C in a rotator set at 250 rpm and then the plasmids expressed in the individual clones were purified using the PureYield Plasmid Miniprep system (Promega, Cat. A2393) according to manufacturer's instructions. The plasmid concentrations and purity were quantitated using the NanoDrop 2000c.

The insertion of the RBD genes into the plasmid pPicZalphaA or pPicZB were confirmed by restriction analysis using 500 ng of purified plasmid and the XhoI (5 units) and XbaI (5 units) restriction enzymes in Cutsmart 1X buffer (1X , NEB) in a volume of 20 μl. The reactions were heated at 70^{o}C for 20 min, subjected to electrophoreseis on a 1.2% agarose gel in 1X TEA electrophoresis buffer (40 mM Tris, pH 8.0, 20 mM acetate , and 1 mM EDTA), and the fragment size evaluated by comparison to the DNA size standard, 1KbPlus DNA ladder (Invitrogen).

2.2.4. Sanger sequencing confirmation of positive clones.

Six specific sequencing primers, that target the internal sequence of the Avian RBD version, were designed based on the initial sequence of the avian RBD monomers, taking care to start the primer sequences with a cytosine (C) or a guanine (G) base (one primer started with a T). The sequencing primers were all approximately 20 bases in length and were shown to be unable to self dimerize and form "primer-dimers". In this way, the sequencing primers used were: FWD520: 5'TCAACTCCAACAATTGCCAC3'; RV524: 5'GTGGCAATTGTTGGAGTTGA3'; FWD715: 5'CTTATTTTGAAGGATTGTTCTG3'; Fwd955: 5'GGTGTTAGTTCTG-CTTGTCC3', FWD1399: 5'CCTACCGATCTTGGTCAATG3'; RV1418: 5'CATTGACC-AAGATCGGTAGG3'. Two plasmid-specific sequencing primers, the 5´ AOX1: 5'GACTGGTTCCAATTGACAAGC3' and 3'AOX1: 5'GCAAATGGCATTCTGACATCC3' targeting the 5' and 3' AOX1 Pichia's promoter region were also used. Using these primers, it was possible to determine the sequences of the avian RBD genes in both the sense and antisense DNA directions to identity the fidelity of the amplification across the

whole ligated sequence. No sequencing primers were designed and synthesized for the human RBD cloned versions (**Figure 6**, **Table 3**).

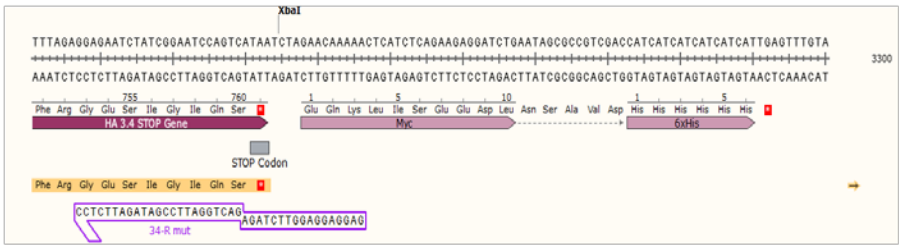
Positive clones with the inserted RBD gene were sequenced using capillary-based fluorescent sequencing on a ABI 3730XL sequencer using a commercial service provider. Individual samples were prepared for sequencing using 250 ng of plasmid containing the RBD gene and 5 pmoles of the appropriate sequencing primer in a total volume of 7.7 μl, the samples were loaded on a 96 well PCR plate, sealed and sent for analysis to the Center for Applied Genomics at the Hospital for Sick Children in Toronto, Canada. The retrieved sequence files from the core facility were analysed using either Sequencher, CLC Workbench, or SnapGene software. The sequences were compared to the sequenced RBD genes originally synthetized by GeneScript, to confirm the removal of the pre-stop codon, correct addition of restriction sites, and lack of nucleotide sequence modifications (insertions, deletion, substitution in nucleotides bases).

Table 2. **Polymerase Chain Reaction (PCR) amplification primers used for modification of the Avian and Human HA-RBD sequences.** The numbers in the primer names denote the Version amplified, with the exception of 3ALL-F and INTRA-F, which were used to amplifiy all the extracellular and intracellular Avian versions, respectively. For each primer 11 base pairs was added at the 5' end of the coding sequence (GAGGAGGAGGT) to promote optimal enzymatic digestion of the XhoI (CTCGAG) or XbaI (TCTAGA) restriction sites (in italics). In the case of intracellular expression a consensus Kozak's sequence (in bold & italic) was inserted for translation initiation, next to a methionine (ATG) start codon (lower capitals). DNA sequences specific for the avian or human HA-RBD versions (in bold).

Primer name	Primer Sequence (5' to 3')	Orientation	Expression Vector
3ALL-F	5'GAGGAGGAGCT*CTCGAG***AAGAGAAGTGTTAAGGGAG** 3'	Forward	pPicZalpha-A
3.4RMUT	5'GAGGAGGAGGT*TCTAGA***GACTGGATTCCGATAGATTCTCC**3'	Reverse	pPicZalpha-A & pPicZ-B
3.6-Rmut	5'GAGGAGGAGGT*TCTAGA***CACTCACACCACTTTGAC**3'	Reverse	pPicZalpha-A & pPicZ-B
3.7Rmut	5'GAGGAGGAGGT*TCTAGA***CTTCTTCTCTGTCTTCTCTTC** 3'	Reverse	pPicZalpha-A & pPicZ-B
3.3-FWD	5'GAGGAGGAGCT*CTCGAG***AAAAGAGGTGTCGCTCCATTGC** 3'	Forward	pPicZalpha-A & pPicZ-B
3.3-R	5'GAGGAGGAGCT*TCTAGA***CCTTTGATGACTTTACTTCTACC** 3'	Reverse	pPicZalpha-A & pPicZ-B
INTRA-F	5'GAGGAGGAGGT*CTCGAG****CGAAACGatgGAG*****AAGAGAAGTGTTAAGGGAG**-3'	Forward	pPicZ-B
F-INTRA3.3		Forward	pPicZ-B

A

Avian HA-RBD V3.4,
3’ PCR Amplification
strategy used for
Extracellular and
Intracellular expression



B

Avian HA-RBD V3.4,
5’ PCR Amplification
strategy used for
Intracellular expression

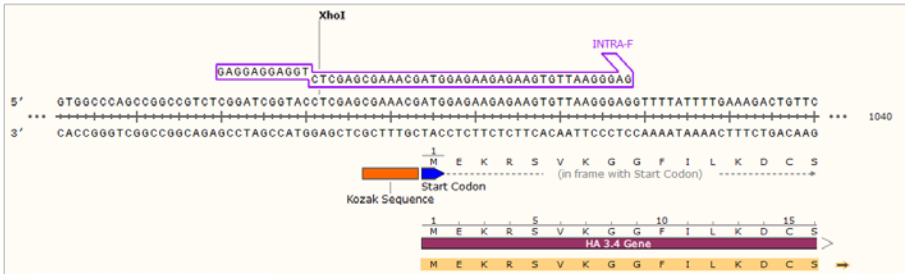


Figure 5. **PCR mutagenesis modifications to enhance Extracellular and Intracellular expression of the original gene design.** In A, a version-specific primer used to remove the premature stop codon originally designed for this study (in this case Avian HA RBD v3.4). The designed primer reconstitutes the open-reading frame to include the c-myc and 6-histidines tag sequences for protein purification and detection. This strategy was applied for the extracellular and intracellular re-cloning of the genes. In B, to increase the level of intracellular expression, it was necessary to design a primer that inserted a XhoI restriction site and Kozak sequence to the 5’ end of the Avian or Human HA-RBD mutated versions.

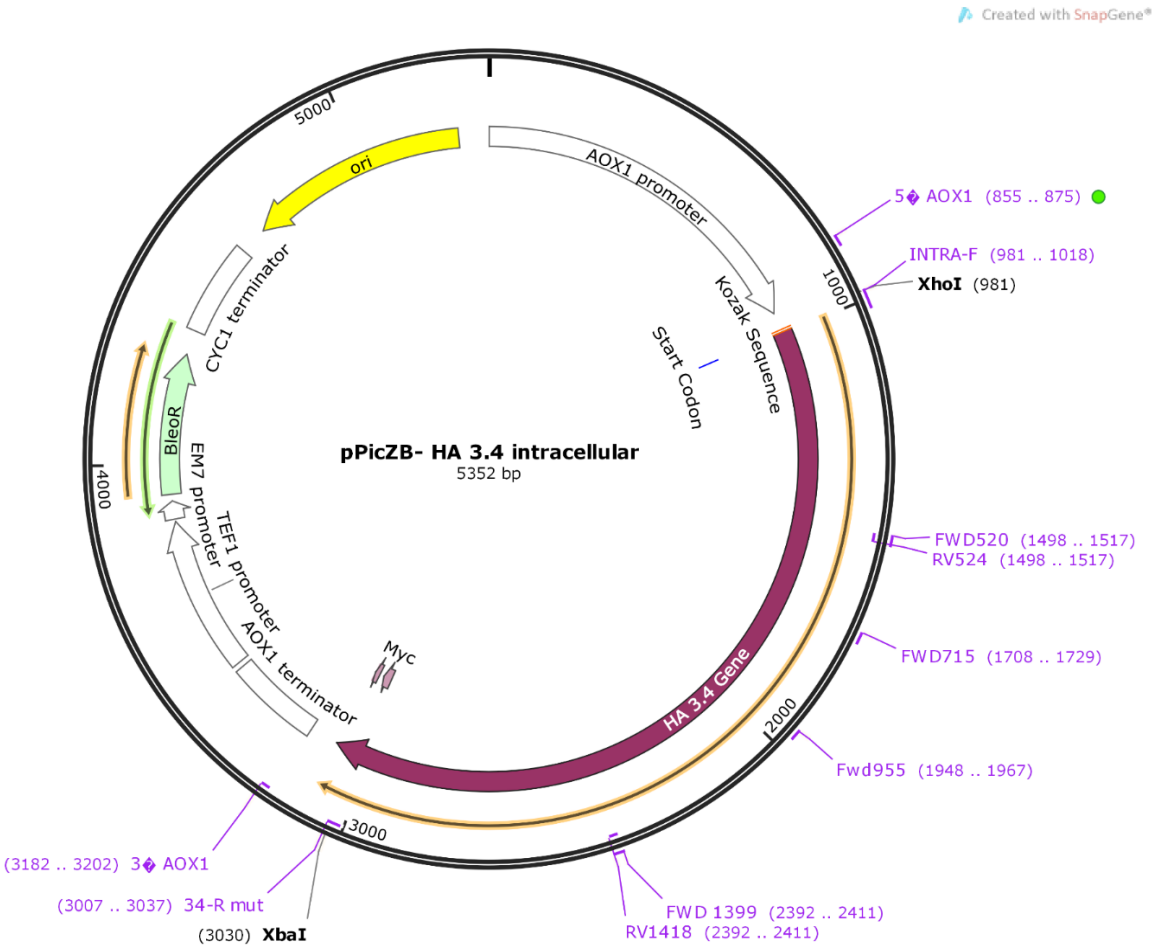


Figure 6. **Schematic localization of the sequencing primers.** The location of the primers used to confirm the correct amplification and cloning of the different types of avian HA RBD genes within the Avian HA RBDs gene located within the pPicZalpha-A or pPicZ-B vectors. A total of 8 primers were used to cover the 2.1 Kb bp length RBD sequence in both DNA strand orientations.

Table 3. List of primers used for sequencing the Avian RBD versions. The number that appears in each primer corresponds to the nucleotide position within the aligned chimeric avian RBD versions.

Primer name	Primer Sequence (5' to 3')	Orientation	Expression Vector
FWD520	TCAACTCCAACAATTGCCAC	Forward	Avian Intracellular and Extracellular
RV524	GTGGCAATTGTTGGAGTTGA	Reverse	Avian Intracellular and Extracellular
Fwd955	GGTGTTAGTTCTGCTTGTCC	Forward	Avian Intracellular and Extracellular
FWD 1399	CCTACCGATCTTGGTCAATG	Forward	Avian Intracellular and Extracellular
RV1418	CATTGACCAAGATCGGTAGG	Reverse	Avian Intracellular and Extracellular
FWD715	CTTATTTTGAAGGATTGTTCTG	Forward	Avian Intracellular and Extracellular
5´ AOX1	GACTGGTTCCAATTGACAAGC	Forward	Avian Intracellular and Extracellular
3´ AOX1	GCAAATGGCATTCTGACATCC	Reverse	Avian Intracellular and Extracellular

2.3. *Pichia pastoris* transformation and antibiotic marker selection

2.3.1. *Pichia pastoris* X-33 strain transformation

Competent *Pichia pastoris* strain X-33 cells were prepared for electroporation using a highly efficiently transformation method (73). Briefly, 5 ml of liquid YPD media culture grown from a single colony were used to inoculate 1 L YPD liquid media (2% peptone, 1% yeast extract, 2% glucose). The cell concentration was monitored until the optical density (OD) of the culture reached 0.6. A total of 1x10^8 cells were harvested by centrifugation and treated in 8 ml of electrochemical competent solution (100 mM LiAc, 10 mM DTT, 0.6 M Sorbitol, 10 mM Tris-HCl, pH 7.5). The suspension was centrifuged at 1500x g for 4 min, and sequentially washed in 1 M Sorbitol three times. The cells were resuspendend to a total volume of 80 μl in electrochemical competent solution and used immediately for electrotransformation. A total of 1 μg of pPicZalphaA-RBD plasmid, which had been linearized by treatment with the SacI restriction enzyme, was incubated on ice in an electro-cuvette containing 80 μL of electrochemical competent *Pinchia pastoris* X-33 cells. The electro-transformation parameters consisted of a single pulse of 1.5 kV, 200 Ω, 25 μF, using a BioRad Gene Pulser system, followed immediately by incubation in ice-cold 1 M Sorbitol solution for 1 h. Following incubation, 20 - 200 μL of the recovered cells were plated on YPDS-agar plates (2% peptone, 1% yeast extract,2% glucose, 1.2 M sorbitol) containing different concentrations of the selective antibiotic Zeocin at 200 μg/mL, 500 μg/mL, 1000 μg/mL and 1500 μg/mL.

2.3.2. Antibiotic selection

The transfected *Pichia pastoris* colonies that grew on solid YPDS-Agar media plates containing zeocin concentrations ranging from 200-1500 μg/mL were selected

(high levels of antibiotic were used in order to select zeocin hyperresistant *P. pastoris* colonies). Between 10-20 resistant *Pichia pastoris* colonies were selected 3-4 days after plating on selective media, and transferred by “toothpick” replating to selective YPD solid media and then transferred to 96-well tissue culture plates for selection in liquid culture to reconfirm hyper-resistance to 1500 µL/mL zeocin antibiotic. The clones grown in 96-wells plates were reselected over three rounds of growth: each clone in the 96-well plate was diluted 1/10th after 24 h of incubation at 260 rpm into a new 96-well plate (Zeocin 200 µg/mL) for each round. After the third selection 5 µl of each clone was transferred to three different 96-well plates with different zeocin concentrations ranging from 500 µg/mL, 1000 µg/mL, and 1500 µg/mL.

2.2.3. Short-term and long-term storage of *Pichia pastoris* transformants

Hyper-resistant selected colonies were plated individually on YPD solid media for short term storage. For long-term storage the clones were cultured in 10 ml YPD liquid media cells harvested by centrifugation at 1500 rpm per 5 min, re-suspended in 1 ml YPD supplemented with 20% glycerol (final concentration), and stored at -80^{o}C.

2.4. Protein expression analysis

2.4.1. *Pichia pastoris* expression growth conditions and media

Zeocin-selected clones were cultured in 200 ml BMGY (1% Bacto Extract, 2% yeast extract, 1% glycerol, 0.005% Biotine, and 2.17% yeast nitrogen base) liquid media until they reached an O.D. between 2 - 6. The cells were harvested by centrifugation and grown in BMMY liquid media containing 1% methanol, to induce recombinant

protein expression, to a final O.D. = 1. The culture was incubated for 48 h and then the cells were harvested by centrifugation at 3500 x g for 30 min. The proteins in the supernatant were subjected to analysis by electrophoresis on polyacrylamide gels containing SDS (SDS-PAGE) and purified using Metal ion Affinity Chromatography (IMAC), gel slice isolation, size exclusion chromatography or anion Exchange Chromatography (AEX) depending of the analysis performed.

2.4.2. Clarification and Ultrafiltration of media supernatant from induced *Pichia pastoris* cultures

Prior to IMAC, the 1 L culture supernatant from the induced *Pichia pastoris* was microfiltered using a 0.2 μm membrane (PALL), was and then ultrafiltered using the QuixStand BenchTop System equipment and a Xampler Filtration Cartridge with a molecular cutoff of 10 KDa. The supernatant was concentrated to 1/10th of the original volume and then diafiltrated/buffer exchanged with PBS, pH7.4 to allow a further decrease in its volume to ~1/20th to facilitate downstream purification.

2.4.3. Anion-Exchange chromatography for the non-6His tagged HA RBD (premature-Stop) proteins.

An anion exchange DEAE (Di-Eethyl-AminoEthyl) resin was used in the first attempt to isolate the full-length, native RBD recombinant proteins. An isolation protocol described for hemagglutinin molecules was used where the supernatants prepared from the cultures induced in 1% methanol for 48 h were concentrated to 1/10th (as described in 2.8.2) and loaded on a 10 ml DEAE-Sepharose column by gravity flow, washed with 2 volumes of phosphate Buffer, pH 7.4 and eluted using a linear gradient of 0 - 1 M NaCl

in phosphate buffer, pH 7.4. The eluted fractions were collected and analyzed by SDS-PAGE, mass-spectroscopy, and immunoblot analysis using polyclonal antibodies In the case of Avian influenza, Ferret (H7N2) antisera to Influenza A virus A/Turkey/Virginia/202 (H7N2)-PR8-IBCDC-5 (IRR FR-1088), Ferret antisera (H5N1) to Influenza A virus A/Chicken/Yunna/1251/2003 was used. In the case of Human Influenza: ferret antisera (H1N1) to influenza A virus A/California/07/2009 (H1N1) pdm09 (IRR FR-359), WHO supplemental antiserum Influenza A (H3N2)v Reference Ferret Antiserum (IRR FR-1000), ferret antiserum (B) to influenza B virus B/Texas/06/2011 (Yamagata Lineage) was used.

2.4.4. Size-Exclusion Chromatography for the DEAE-purified protein fractions

After DEAE-chromatography, the eluted fractions were examined for appropriately sized proteins using Mass Spectrometry. Fractions containing putative protein bands where further separated on an AKTA system for protein purification using a DEAE column. The fractions were loaded onto the column at a flow speed of 1 mL/min, the columns washed with 2 column volumes of phosphate buffer, pH 7.4 at 1 mL/min, and then eluted over a linear gradient of 0 - 1 M NaCl in phosphate buffer, pH 7.4. The fractions corresponding to the protein peaks were monitored using UV absorbance and were then collected and pooled for Mass Spectroscopic analysis and Immunodetection with polyclonal antibodies.

2.4.5. Tandem Mass-Spectrometry of DEAE- and Size-Exclusion purified non-6-HIS tagged HA-RBD peptides

For the tandem Mass-Spec analysis, protein bands in the range of 72-120 KDa in size were excised directly from the SDS-PAGE gels using a sterile scalpel in a biosafety cabinet in order to avoid an external source of protein contamination for the analysis. The gel slice containing the band of interest was stored in 1% acetic acid prepared using ultrapure water (Sigma). Proteomics analysis was performed at the OHRI Proteomics Core Facility (Ottawa, Canada).

2.4.6. Collaborative Methodologies

As stated by the Core Facility guidelines, proteins were digested in-gel using trypsin (Promega) according to the method of Shevchenko (74). The resulting peptide extracts were concentrated using a vacufuge (Eppendorf), purified and desalted using a ZipTip (Millipore), and resuspended in 1% formic acid (Fisher). Peptides were analyzed by LC-MS/MS (liquid chromatography – tandem mass spectrometry) on a system comprised of an UltiMate 3000 RSLC nano-HPLC, LTQ Orbitrap XL hybrid mass spectrometer and nanospray ionization source (Thermo Scientific). The system was controlled by Xcalibur software, version 2.0.7 (Thermo Scientific).

Peptides were loaded by autosampler onto a trap column (C18 CapTrap; Michrom, USA) in 3% acetonitrile, 0.1% formic acid at a flow rate of 15 μl/min for 5 min. Peptides were eluted over 60 minute on a gradient of 3% - 45% acetonitrile at a flow rate of 300 nl/min through a 10-cm long column with integrated emitter tip (Picofrit PF360-75-15-N-5 from New Objective packed with Zorbax SB-C18, 5 μm from Agilent),

and nanosprayed into the mass spectrometer. The nanoflow HPLC solvents contained 0.1% formic acid and 5% DMSO (75). MS scans were acquired in FTMS mode at a resolution setting of 60,000. MS^2 scans were acquired in ion trap CID mode using data-dependent acquisition of the top 5 ions from each MS scan.

MASCOT software version 2.5 (Matrix Science) was used to infer peptide and protein identities from the mass spectra data. The observed MS/MS spectra were matched against sequences from SwissProt version 2013_05, a database of common contaminants, and database of customer-supplied sequences. Mass tolerance parameters were set to MS ±10 ppm and MS/MS ±0.6 Da. Enzyme specificity was set to 'Trypsin' with $\leq$2 miscuts. Oxidation of methionine, protein N-terminal acetylation, pyrocarbamidomethlyation of N-terminal cysteine, and conversion of glutamine to pyroglutamate were allowed as variable modifications. Carbamidomethylation of cysteine was set as a fixed modification.

2.4.7. Immobilized Metal Affinity Chromatography (IMAC) for His-Tagged RBD proteins

For IMAC purification, the culture supernatant from methanol-induced *Pichia pastoris,* zeocin-hyperrresistant clones was filtered through 0.2 µm filters (Stericup, Nalgene), concentrated using the Quixstand ultrafiltration device and filtered using a 10,000 molecular weight cut off filter (General Electric), and then buffer exchanged for PBS, pH 7.4. The sample was then mixed with 1 ml of Ni-NTA agarose buffer-equilibrated resin in a 50 ml centrifuge tube and mixed for 1.5 hours at room temperature and 10 rpm.

The Ni-NTA resin was centrifuged at 1500 x r.p.m. for 5 minutes to settle the beads and the resin slurry transferred into a 10 ml chromatographic column. The sample contained in the agarose beads was washed with 7 bead-volumes of Washing Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) in order to remove protein impurities and weakly interacting (histidines-containing proteins) and unwanted proteins. The washed beads were immediately eluted with 3 ml of Elution Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0) and 0.5 ml fractions were collected in 1.5 ml centrifuge tubes. Samples of each fraction were analyzed for the presence of the His-tagged protein following electrophoresis on 10% PAGE containing SDS 10% gels and immunoblot analysis using an anti-Histidine mouse monoclonal antibody (Roche Ltd). After confirming the presence of the appropriately sized reactive proteins(s), the positive fractions were pooled, concentrated and buffer exchanged with PBS, pH 7.5, using a mini ultrafiltration device (10,000 MWCO, Millex-Millipore) following the manufacture's instructions. Protein concentration of the purified Avian or Human RBDs, were determined either by a BCA protein assay kit (Pierce, Cat No.23225) or by UV-spectrophotometry reading at 280 nm using a NanoDrop 2000s.

2.4.8. Sodium Dodecyl-Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) Protein Analysis.

Protein samples were analysed on 10% polyacrylamide gels containing SDS (SDS-PAGE), and subjected to electrophoresis at 120 Volts for 1.5-2 h or until full separation of the protein size standard (Blueye Protein Ladder, Genedirex) was present.

2.4.9. Immunodetection by immunoblot analysis

The first strategy for immunodetection following protein purification of the HA RBD that did not contain a 6-His tag, employed the use of a mix of primary ferret-polyclonal antibodies raised against the Avian Influenza virus AH5N1 and AH7N3 or ferret-polyclonal antibodies raised against human influenza virus A/H1N1 and B influenza, obtained from the CDC-Influenza Research Repository (address). A goat anti-ferret-HRP was used as the secondary detection antibody. The mutated versions HA-RBDmut-V3.3, HA-RBDmut-V3.4 and HA-RBDmut-V3.6, were probed using only an anti-6His-HRP mouse monoclonal antibody (Roche diagnostics, Catalogue No.11965085001). Briefly, HA-RBD purified by IMAC were boiled for 5 minutes in SDS Loading Buffer (0.25 % bromophenol blue; 50 mM dithiothreitol; 20% glycerol; 4% sodium dodecyl sulfate; 0.125 M Tris-Cl, pH 6.8). The denatured RBDs samples were loaded on a 5% stacking gel/ 7%-running SDS-PAGE gel to facilitate protein separation as above. The gels were transferred to PVDF membranes (Biorad) that were previously activated by soaking in absolute methanol. Ice-cold Transfer Buffer/SDS (25 mM Tris-Cl, 192 mM glycine, 20% v/v methanol, 5% SDS) was used to facilitate the movement of high molecular size polypeptides out of the gel. The current applied for the transfer was determined empirically, with the best results obtained using 25 V for 18 min using a Semi-Dry transfer cell (Biorad). A prestained protein standard (Blueye, Genedirex) helped to visually assess the optimal transfer. The PVDF immobilized RBDs samples were blocked by incubation in 1X PBS/0.5%Tween20/5%skim milk (PBST/M), pH 7.4, and washed with PBST three times for 5 minutes each. The blots were then probed with anti-6 His mouse monoclonal antibody-coupled to horseradish peroxidase (titre of

1:500) in PBST/1% skim milk, and incubated for 2 h with slight agitation at room temperature. After incubation, the blots were washed three times with PBST for 5 minutes each. To detect the bound antibodies, the blots were incubated in 10 ml of 1-Step™ Ultra TMB-Blotting Solution (Cat No.37574, ThermoFisher) following the manufacturer instructions to cover the membrane area to fully develope the enzymatic reaction. The evolved chromogenic reaction was detected using a digital camera.

2.4.10. Deglycosylation Assays

For the identification of the RBD glycosylation patterns in the recovered His-Tagged HA-RBD mutV3.4, HA RBD mutV3.6, and HA RBD mutV3.7 proteins, 1-20 μg of sample protein were combined with 1 µl of 10X glycoprotein denaturing buffer to a total volume of 10 µl (with water)total reaction volume. The protein samples were denatured by heating at 100 ^{o}C for 10 min. To the denatured samples, 10 μl was added containing, 2 μl 10X Glycobuffer, 2 μl endoglycosidase EndoHf (2000 units total, NEBiolabs), and H_2O (if necessary) to 10 μl, both solutions were mixed gently and incubated at 37 ^{o}C for 1 h. The de-glycosylation protein pattern was analysed using a 10%SDS-PAGE gel, transferred to a PVDF (Biorad) membrane for western blot detection using anti-hexahistidine-HRP mouse monoclonal antibody (Roche). Avian or Human RBDs protein concentrations were determined either by a BCA protein assay kit (Pierce, Cat No.23225) or by the UV-spectrophotometry reading at 260 nm using the NanoDrop equipment.

2.4.11. Hemmaglutination assays in Guinea Pig Red Blood Cells

Serial dilutions composed of 50 μl of purified RBD proteins in PBS pH 7.5 , were incubated with 50 μl of Guinea pig red blood cells (RBC) 0.5 % in PBS pH 7.4, for one h. The amount of hemagglutination was recorded and compared to positive controls consisting of Control antigen (H1) pdm09-A/California/07/2009-NYMC X-179A (Catalogue No. FR 1184, IRC) and a positive control consisting of the trivalent Fluviral vaccine version 2013-2014 (GlaxoSmith&Kline). The negative control consisted of PBS, pH7.5, only. The scores were based on the guidelines and lab procedures suggested by the National Veterinary Services Laboratory, USA (76).

2.5 Bioinformatics tools

2.5.1. Protein homology/analogY Recognition Engine version2.0 (Phyre2) as a tool for protein folding prediction.

The linear sequence for each monomer protein contained in the avian and in the human RBDs versions were analysed individually using the Phyre2 (72) server under “normal” mode modelling, and for the multimeric gene design uder “intensive” mode. Protein DataBank (*.pdb) files were generated to predict and analyze protein structure based on sequence homology with PSI-Blast database. The purpose of this 3-D modelled protein analysis was merely exploratory, but helped to visualize the predicted glycosylated sites in the HA RBD sequences and other post-translated modifications observed in experimental results.

2.5.2. glycosylation modelling of proteins using Gly-Pro.

Based on the difference between the experimentally observed N-glycosylated and N-deglycosylated form of the human and avian RBDs, the software GlyPro (Glycosciences) was used in order to create *in silico* models of potential N-glycosylation sites based on the presence of the sequence motif, N-X-S/T/C (X not proline). The putative protein glycosylation sites were modeled on the human and avian modelled RBD PDB files created with Phyre2 software. The predicted sites and carbohydrate superimposition were reflected over the surface of the 3D protein structure.

2.5.3. The use of ProP V1.0 as a tool for protease signal sequence prediction

In order to explain certain experimental results suggesting the presence of an internal cleavage site during processing within the Human HA RBD version, the predictor software ProP V1.0 (77) was employed. Protease cleavage (PC) scores above the default predictor threshold values were only taken into account to suggest internal endopeptidase cleavage sites within the Avian and Human HA RBD versions. Using the same software, a series of hypothetical single residues modifications were performed manually and retrieved for software analysis to evaluate the best mutagenesis substitution that was able to abolish the cleavage signal predictor score to suggest mutagenesis modifications of the multimeric RBD gene.

3.0 RESULTS

3.1. Antigen design

3.1.1. Antigen selection of human HA RBD Subunits for the HA-RBD multimeric versions.

Three HA RBD subtypes comprising the 2013-2014 vaccine composition (1) were selected from full Hemagglutinins (HA0) sequences directly retrieved from GeneBank (**Figure 7**) and analysed using Clustal W/UCSD Chimera. Published Human globular RBD1 (GenBank: ACQ99608.1) from A/California/7/2009 (H1N1)pdm09-like (GenBank: YP_009118626.1), was used as a backbone sequence for the polypeptide selection of the RBD2 (A/Texas/50/2012(H3N2) (GenBank: AIE52525.1) ((**Figure 9**) and RBD3 (B/Brisbane/33/2008) (GenBank: ACN29387.1) (**Figure 11**) subunits. As a consensus, the RBD subunits avoided a basic patch region found in the HA1 domain and the fusion peptide-transmembrane helix sequences that form part of the Hemagglutin HA2 domain, as previously reported for RBD1 (2,3) (**Table 4**). The sequences subunits selected were analyzed individually in Phyre2 server. All sequences predicted the lectin globular structural domain for RBD1, RBD2 and RBD3 and their modelled structures superimposed correctly to extant HA's X-Ray diffraction data for A/California/7/2009 (H1N1) pdm09-like (PDB 4M4Y) (**Figure 8**), to A/Victoria/361/2011 (H3N2) (PDB 4WE9) (**Figure 10**) and (B/Brisbane/33/2008) (Influenza B) (PBD 4FQK) (**Figure 12**), respectively.

Antigen selection for Human HA-RBD Subunit 1

Hemagglutinin Influenza A virus (A/California/07/2009(H1N1)

```
ORIGIN
        1 mkailvvlly tfatanadtl cigyhannst dtvdtvlekn vtvthsvnll edkhngklck
       61 lrgvaplhlg kcniagwilg npeceslsta sswsyivetp ssdngtcypg dfidyeelre
      121 qlssvssfer feifpktssw pnhdsnkgvt aacphagaks fyknliwlvk kgnsypklsk
      181 syindkgkev lvlwgihhps tsadqqslyq nadayvfvgs srysktfkpe iairpkvrdr
      241 egrmnyywtl vepgdkitfe atgnlvvpry afamernags giiisdtpvh dcnttcqtpk
      301 gaintslpfq nihpitigkc pkyvkstklr latglrnips iqsrglfgai agfieggwtg
      361 mvdgwygyhh qneqgsgyaa dlkstqnaid eitnkvnsvi ekmntqftav gkefnhlekr
      421 ienlnkkvdd gfldiwtyna ellvllener tldyhdsnvk nlyekvrsql knnakeigng
      481 cfefyhkcdn tcmesvkngt ydypkyseea klnreeidgv klestriyqi laiystvass
      541 lvlvvslgai sfwmcsngsl qcrici
//
```

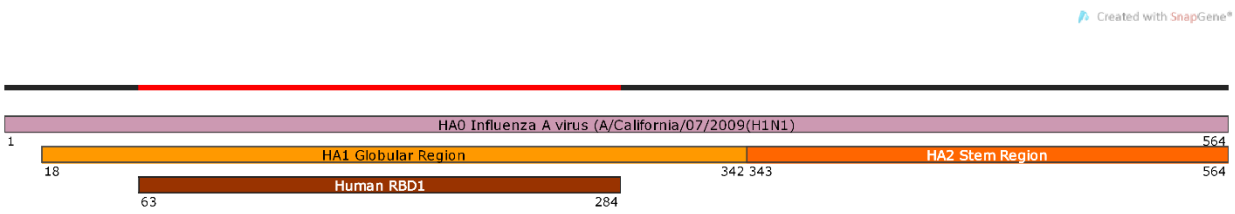


Figure 7. Diagram alignment showing the HA-RBD1 residues selected from the Influenza Hemagglutinin A/California/7/2009 (H1N1) pdm09-like virus. Virus selection was based on the recommendation by WHO for the 2013-2014 trivalent influenza vaccine composition (1). The RBD1 region (in bold letters and as Human RBD1 in diagram) is part of HA1 domain but lacks its charged patch region and additionally excludes the fusion peptide and transmembrane domain and all the residues of the HA2 stem region (4).

Modelling and Structure comparison for Human HA-RBD Subunit1

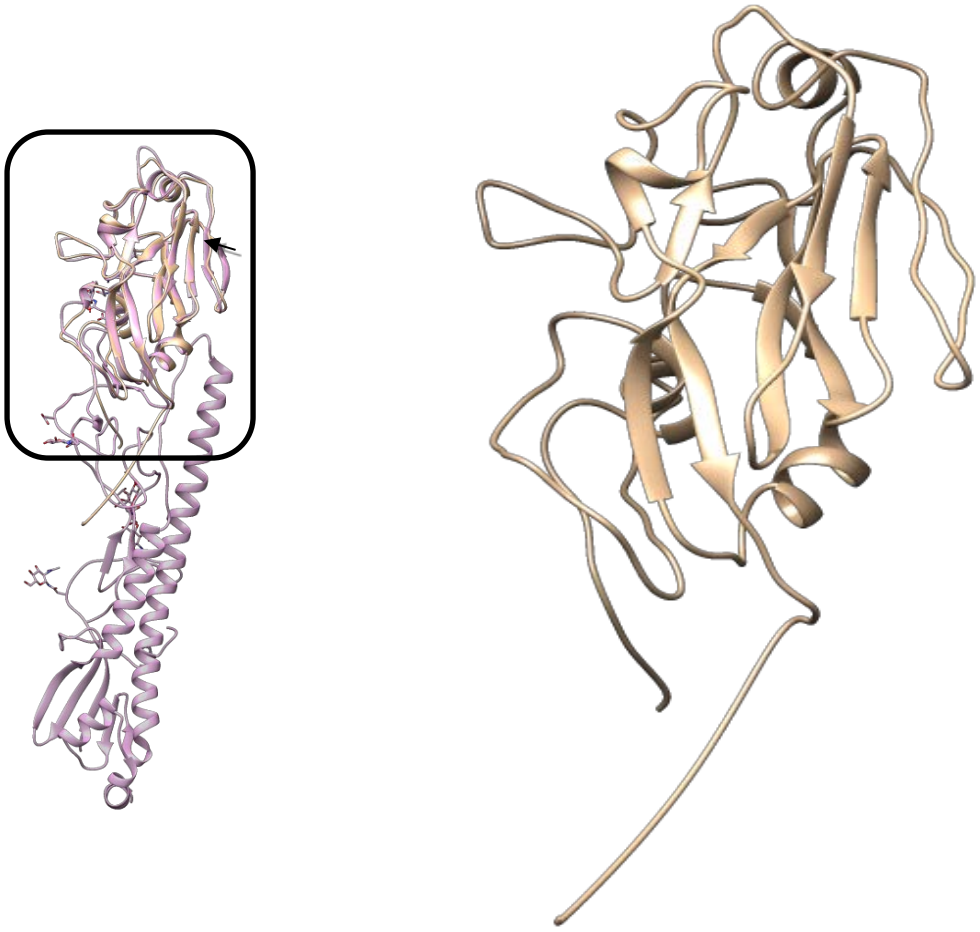


Figure 8. Structural model of the Human HA RBD subunit 1. The PDB file of the globular head comprising the human HA RBD Subunit 1(green) was generated by the Phyre2 server (5) and superimposed over the determined crystal structure map of the Hemagglutinin HA1 (in cyan) for the virus A/California/07/2009 (H1N1) pandemic virus (PDB file 4M4Y) for comparison and illustration of the region selected for antigen expression in the present study (black box). The arrow indicates the receptor binding site for the HA1 molecule. The graphs were generated using Chimera UCSF software, and the crystal structure PDB file retrieved from the RCSB Protein Data Bank website.

Antigen selection for Human HA-RBD Subunit 2

Influenza A virus (A/Texas/50/2012(H3N2))

```
ORIGIN
        1 mktiialsyi lclvfaqklp gndnstatlc lghhavpngt ivktitndri evtnatelvq
       61 nssigeicds phqildgenc tlidallgdp qcdgfqnkkw dlfverskay sncypydvpd
      121 yaslrslvas sgtlefnnes fnwngvtqng tssacirrsn nsffsrlnwl thlnfkypal
      181 nvtmpnneqf dklyiwgvhh pgtdkdqifl yaqpsgritv stkrsqqavi pnigsrprir
      241 nipsrisiyw tivkpgdill instgnliap rgyfkirsgk ssimrsdapi gkcksecitp
      301 ngsipndkpf qnvnrityga cpryvkqstl klatgmrnvp ekqtrgifga iagfiengwe
      361 gmvdgwygfr hqnsegrgqa adlkstqaai dqingklnrl igktnekfhq iekefseveg
      421 riqdlekyve dtkidlwsyn aellvalenq htidltdsem nklfektkkq lrenaedmgn
      481 gcfkiyhkcd nacigsirng tydhdvyrde alnnrfqikg velksgykdw ilwisfaisc
      541 fllcvallgf imwacqkgni rcnici
//
```

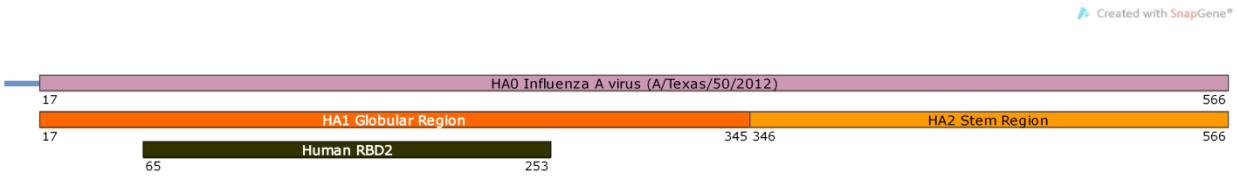


Figure 9. Sequence alignment showing the HA-RBDs residues selected from the Influenza Hemagglutinin of Influenza A virus (A/Texas/50/2012(H3N2)). Virus selection was based on the recommendation by WHO for the 2013-2014 trivalent influenza vaccine composition (1). The RBD2 region (in bold letters and as Human RBD2 in diagram) is part of the HA1 globular region and excludes the fusion peptide and transmembrane domain and all the residues of the HA2 stem region (6).

Modelling and Structure comparison for Human HA-RBD Subunit 2

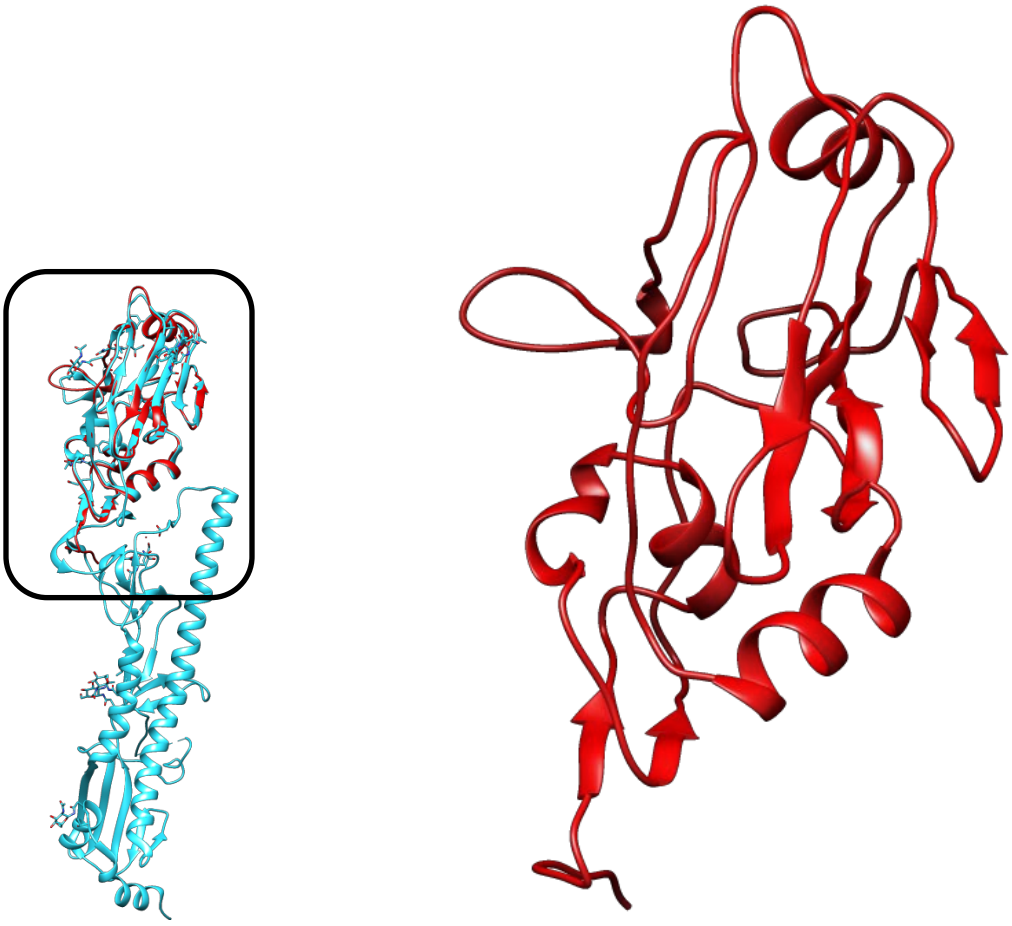


Figure 10. Structural model of the Human HA RBD subunit 2. The PDB file of the globular head comprising the human HA RBD Subunit 2 (red) was generated by the Phyre2 server (5) and superimposed over the determined crystal structure map of the Hemagglutinin HA3 (in cyan) of the virus A/Victoria/361/2011 (PDB file 4WE9) for comparison and illustration of the region selected for antigen expression in the present study (black box). The graphs were generated using Chimera UCSF software, and the crystal structure PDB file was retrieved from the RCSB Protein Data Bank website.

Antigen selection for Human HA-RBD Subunit 3

Influenza B virus (B/Brisbane/33/2008)

```
ORIGIN
        1 mkaiivilmv vtsnadrict gitssnsphv vktatqgevn vtgviplttt ptkshfanlk
       61 gte**trgklcp kclnctdldv algrpkctgk ipsarvsilh evrpvtsgcf pimhdrtkir**
      121 **qlpnllrgye hirlsthnvi naenapggpy kigtsgscpn itngngffat mawavpkndk**
      181 **nktatnplti evpyictege dqitvwgfhs dsetqmakly gdskpqkfts sangvtthyv**
      241 **sqiggfpnqt edgglpqsgr ivvdymvqks gktgtityqr gillpqkvwc asgrskvikg**
      301 slpligeadc lhekygglnk skpyytgeha kaigncpiwv ktplklangt kyrppakllk
      361 ergffgaiag fleggwegmi agwhgytshg ahgvavaadl kstqeainki tknlnslsel
      421 evknlqrlsg amdelhneil eldekvddlr adtissqiel avllsnegii nsedehllal
      481 erklkkmlgp saveigngcf etkhkcnqtc ldriaagtfd agefslptfd slnitaasln
      541 ddgldnhtil lyystaassl avtlmiaifv vymvsrdnvs csicl
//
```

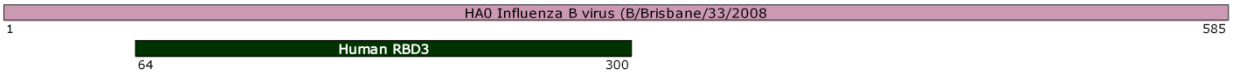


Figure 11. Sequence alignment showing the HA-RBD selected from the Influenza Hemagglutinin of Influenza B virus (B/Brisbane/33/2008). Virus selection was based on the recommendation by WHO for the 2013-2014 trivalent influenza vaccine composition (1). RBD region (in bold letters and as Human RBD3 in diagram) is part of the HA2 globular region and excludes the fusion peptide and transmembrane domain and all the residues of the HA2 stem region (7).

Modelling and Structure comparison for Human HA-RBD Subunit 3

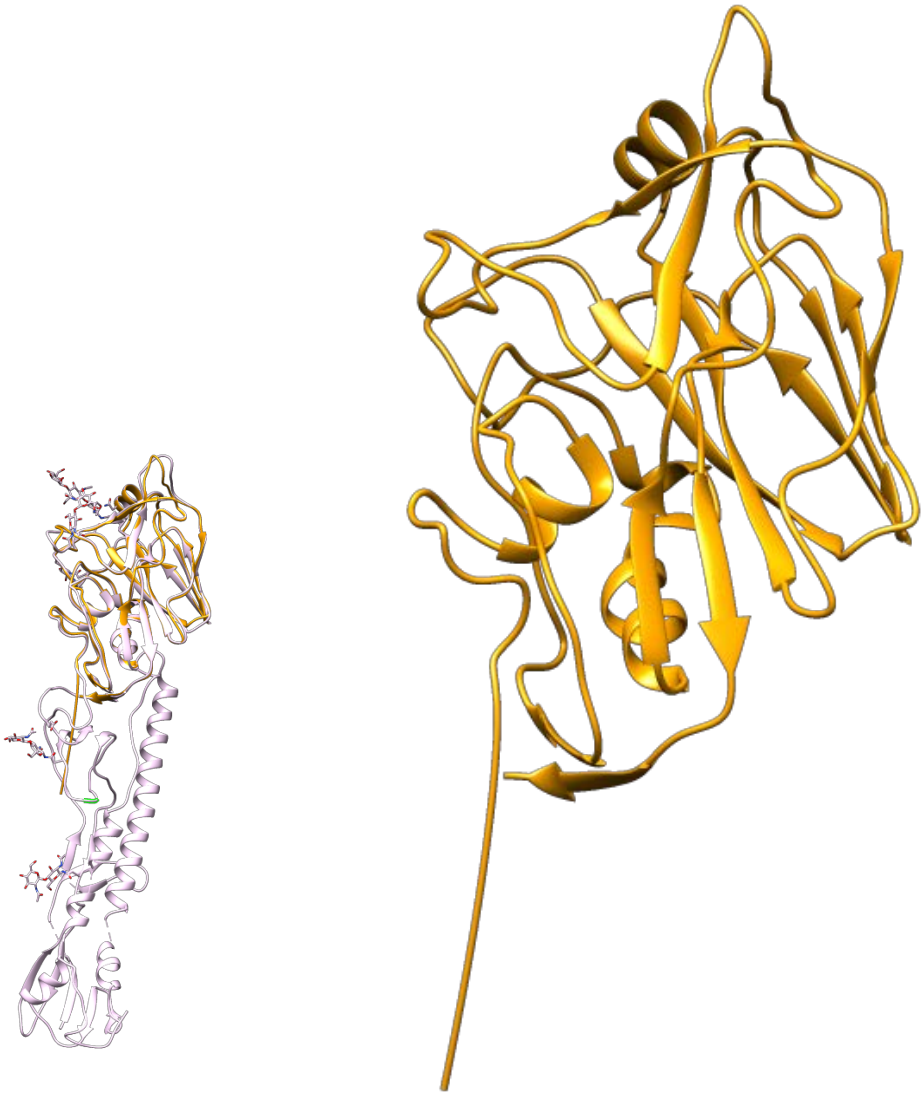


Figure 12. Structural model of the Human Influenza B, HA-RBD subunit 3. The PDB file of the globular head comprising the human HA-RBD Subunit 3 (orange) was generated by the Phyre2 server (5) and superimposed over the determined crystal structure map of the Influenza B Hemagglutinin (in cyan) of the virus B/Brisbane/60/2008 (PDB file 4FQK) for comparison and illustration of the region selected for antigen expression in the present study (black box)(8). The graphs were generated using Chimera UCSF software, and the crystal structure PDB file was retrieved from the RCSB Protein Data Bank website.

Table 4. Protein Sequence alignment for the Human HA's RBD Subunits used to construct the Human multimeric gene versions. Using the previously reported RBD1 (A/California/07/2009(H1N1) sequence (2,3) as a template, RBD2 (A/Texas/50/2012(H3N2), RBD3 (B/Brisbane/33/2008), were aligned for comparison to IV A/California/07/2009(H1N1) Hemagglutinin used as a template to determine the border sequences of the lectin domain and was based on the conserved residues share among all the sequences used as a backbone scaffold guidance, their predicted globular folding generated by Phyre2 server presented in **Figures 8,10,12**. (*), asterisk indicated fully conserved residue; (:), semicolon indicates conservation among groups of residues with strongly similar properties; (.), period indicates residues conservation among weakly similar properties; (highlighted in grey), sequence corresponding to the fusion peptide present in HA2 domain; (residues in closed boxes), represent a charged patch region in the HA1 domain formed when HA1-HA2 folds natively, it was excluded from the final sequence. The alignment was generated using ClustalW program from the Chimera UCSF software.

```
CLUSTAL W ALN saved from UCSF Chimera/MultAlignViewer

HA_[A/California/07/2009(H1N1]      MKAILVV---LLY------TFATANADTLCIGYHANNSTDTVDTVLEKNV 41
HA_[A/Texas/50/2012(H3N2]           MKTIIALSYILCLVFAQKLPGNDNSTATLCLGHHAVPNGTIVKTITNDRI 50
HA_[B/Brisbane/33/2008]             MKAIIVILM-----------VVTSNADRICTGITSSNSPHVVKTATQGEV 39
                                    **:*:.:                .:  :* *  :  .   *.*  : .:

HA_[A/California/07/2009(H1N1]      TVTHSVNLLEDKHNGKLCKLRG-----VAPLHLGKCNIAGWILGNPECES 86
HA_[A/Texas/50/2012(H3N2]           EVTNATELVQNSSIGEICDSPH------QILDGENCTLIDALLGDPQCDG 94
HA_[B/Brisbane/33/2008]             NVTGVIPLTTTPTKSHFANLKGTETRGKLCPKCLNCTDLDVALGRPKCTG 89
                                     **    *      ..:..            .  :*.  .  ** *:* .

HA_[A/California/07/2009(H1N1]      LSTASSWSYIVETPSSDNGTCYPGDFIDYEELREQLSSVSSFERFEIF-- 134
HA_[A/Texas/50/2012(H3N2]           FQN-KKWDLFVER-SKAYSNCYPYDVPDYASLRSLVASSGTL---EFN-- 137
HA_[B/Brisbane/33/2008]             KIPSARVSILHEV-RPVTSGCFPIMH-DRTKIRQLPNLLRGYEHIRLSTH 137
                                           . : *      . *:*    *  .:*.           .:

HA_[A/California/07/2009(H1N1]      --PKTSSWPNHDSNKGVTAACPHAG-AKSFYKNLIWLVKKGNSYPK---- 177
HA_[A/Texas/50/2012(H3N2]           --NESFNWNG-VTQNGTSSACIRRS-NNSFFSRLNWLTHLNFKYPA---- 179
HA_[B/Brisbane/33/2008]             NVINAENAPGGPYKIGTSGSCPNITNGNGFFATMAWAVPKNDKNKTATNP 187
                                       :: .  .   : *.:.:* .    :.*:  : * .  . .

HA_[A/California/07/2009(H1N1]      --LSKSYINDKGKEVLVLWGIHHPSTSADQQSLYQNADAYVFVGS-SRYS 224
HA_[A/Texas/50/2012(H3N2]           --LNVTMPNNEQFDKLYIWGVHHPVTDKDQIFLYAQPSGRITVST-KRSQ 226
HA_[B/Brisbane/33/2008]             LTIEVPYICTEGEDQITVWGFHSDSETQM-AKLYGDSKPQKFTSSANGVT 236
                                      :. .    :  : : :**.*          ** :..    ..: .

HA_[A/California/07/2009(H1N1]      KKFKPEIAIRPKVR-----DQEGRMNYYWTLVEPGDKITFEATGNLVVPR 269
HA_[A/Texas/50/2012(H3N2]           QAVIPNIGFRPRIR-----NIPSRISIYWTIVKPGDILLINSTGNLIAPR 271
HA_[B/Brisbane/33/2008]             THYVSQIGGFPNQTEDGGLPQSGRIVVDYMVQKSGKTGTITYQRGILLPQ 286
                                        .:*.  *.          .*:   : : :.*.   :    .:: *:

HA_[A/California/07/2009(H1N1]      YAFAMERNAGSGIIISDTPVHDCNTTCQTPKGAINTSLPF-QNIHPITIG 318
HA_[A/Texas/50/2012(H3N2]           GYFKIR-SGKSSIMRSDAPIGKCKSECITPNGSIPNDKPF-QNVNRITYG 319
HA_[B/Brisbane/33/2008]             KVWCA--SGRSKVIKGSLPLIGEADCLHEKYGGLNKSKPYYTGEHAKAIG 334
                                      :    .. * :: .. *:           *.: .. *:  . :  : *

HA_[A/California/07/2009(H1N1]      KCPKYVKSTKLRLATGLRNIPS---IQSRGLFGAIAGFIEGGWTGMVDGW 365
HA_[A/Texas/50/2012(H3N2]           ACPRYVKQSTLKLATGMRNVPE---KQTRGIFGAIAGFIENGWEGMVDGW 366
HA_[B/Brisbane/33/2008]             NCPIWVKT-PLKLANGTKYRPPAKLLKERGFFGAIAGFLEGGWEGMIAGW 383
                                     ** :**   *:**.* :  *     : **:*******:*.** **: **

HA_[A/California/07/2009(H1N1]      YGYHHQNEQGSGYAADLKSTQNAIDEITNKVNSVIEKMNTQFTAVGKEFN 415
HA_[A/Texas/50/2012(H3N2]           YGFRHQNSEGRGQAADLKSTQAAIDQINGKLNRLIGKTNEKFHQIEKEFS 416
HA_[B/Brisbane/33/2008]             HGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMD 433
                                    :*:  :. .* . ******** **::*. ::* :      ::  :   :.

HA_[A/California/07/2009(H1N1]      HLEKRIENLNKKVDDGFLDIWTYNAELLVLLENERTLDYHDSNVKNLYEK 465
HA_[A/Texas/50/2012(H3N2]           EVEGRIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEMNKLFEK 466
HA_[B/Brisbane/33/2008]             ELHNEILELDEKVDDLRADTISSQIELAVLLSNEGIINSEDEHLLALERK 483
                                    .:. .* :*:: *:*   *  : : ** * *.*:  ::  *..:  * .*

HA_[A/California/07/2009(H1N1]      VRSQLKNNAKEIGNGCFEFYHKCDNTCMESVKNGTYDYPKYSEEAKLNRE 515
HA_[A/Texas/50/2012(H3N2]           TKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDHDVYRDEALNNRF 516
HA_[B/Brisbane/33/2008]             LKKMLGPSAVEIGNGCFETKHKCNQTCLDRIAAGTFDAGEFSLPTFDS-L 532
                                     :. *  .* ::*****:  ***:::*:  :  **:*   :   :  .

HA_[A/California/07/2009(H1N1]      EIDGVKLE--STRIYQILAIYSTVASSLVLVVSLGAISFWMCSNGSLQCR 563
HA_[A/Texas/50/2012(H3N2]           QIKGVELK--SGYKDWILWIS-FAISCFLLCVALLGFIMWACQKGNIRCN 563
HA_[B/Brisbane/33/2008]             NITAASLNDDGLDNHTILLYYSTAASSLAVTLMIAIFVVYMVSRDNVSCS 582
                                    :* ...*:  .     **     . *.: : : :  : .:  ....: *

HA_[A/California/07/2009(H1N1]      ICI 566
HA_[A/Texas/50/2012(H3N2]           ICI 566
HA_[B/Brisbane/33/2008]             ICL 585
                                    **:
```

3.1.2. Antigen selection of Avian HA-RBD subunits for the HA-RBD multimeric versions

Three Avian Hemagglutinin polypeptide sequences, identified from RT-PCR sequencing of the HA1 domain from Avian virus were circulating among Chicken Farms from México, were provided by the vaccine company Investigación Aplicada S.A. The residues of three Avian HA RBD subtypes were trimmed using as a template the published Human globular RBD1 (GenBank: ACQ99608.1) from A/California/7/2009 (H1N1) pdm09-like (GenBank: YP_009118626.1 (**Figure 7**) and analysed using Clustal W/UCSD Chimera. All NCBI Blast results for the RBD sequences confirmed its Avian origin, best- scored sequences identifies both RBD1 (**Table 4**) and RBD2 (**Table 6**) as high homologue to A/chicken/Hidalgo/1336-06/2006(H5N2) (GenBank: AIN25320.1); and best-scored sequence identifies RBD3 (**Table 8**) as (A/Turkey/Ontario/18/2000(H7-like) (GenBank: AF497552_1). As a consensus, the parameters for RBD subunits selection was based on the closest first and last conserved residues the Avian HA1 sequences shared with the backbone sequence of Human globular RBD1 peptide (2,3) (**Table 11**) used as scaffold. The selected sequences subunits were analyzed individually in Phyre2 server. All sequences predicted the lectin globular structural domain for RBD1 (**Figure 13**), RBD2 (**Figure 15**) and RBD3 (**Figure 17**). The modelled structures superimposed to extant HA's X-Ray diffraction data in A/duck/Egypt/10185SS/2010 (H5N1) (PDB file 5E2Y) (**Figure 13-14**, **Table 5**), in A/Victoria/361/2011 (H3N2) (PDB 5E2Y)(**Figure 15-16**, **Table 7**) and in A/Anhui/1/2013 (PDB file 4R8W) (H7-like) (**Figure 17-18**, **Table 9**), respectively.

Table 5. Antigen selection for Avian HA-RBD Subunit 1 (H5-like) Sequence BLAST for Avian HA-RBD Subunit 1. The HA RBD Subunit 1 sequence was derived from a field virus isolation from chicken farms located in Mexico. The sequences provided by IASA Company, were used in Blast searches of the NCBI database to compare and identify the origin of the Hemagglutinin RBD gene product used in this study. The HA-RBD subunit 1 has a high protein homology to the H5-like hemaglutinin protein. The best score was obtained for the Influenza A virus (A/chicken/Hidalgo/1336-06/2006(H5N2) virus isolate present in the NCBI database.

Hemaglutinin [Influenza A virus (A/chicken/Hidalgo/1336-06/2006(H5N2))

Sequence ID: gi|682021806|gb|AIN25320.1|Length: 564Number of Matches: 1

Score	Expect	Method	Identities	Positives	Gaps
441 bits(1134)	1e-151	Compositional matrix adjust.	209/224(93%)	215/224(95%)	2/224(0%)

```
Query  1    SVKG--GFILKDCSVAGWLLGNPMCDEFLTAPEWSYIVEKDNPSNDLCYPGNFDNYEELK  58
            S+KG    ILKDCSVAGWLLGNPMCDEFL+ PEWSYIVEKDNPSNDLCYPGNFDNYEELK
Sbjct  59   SLKGIKPLILKDCSVAGWLLGNPMCDEFLSVPEWSYIVEKDNPSNDLCYPGNFDNYEELK  118

Query  59   HLMSSTNHMERIQIFPRNSWTNHNASAGVSSACPFNGGSSFFRNVVWLIKKNDVYRTLKR  118
            HLMSSTNH+ERIQIFPRNSW++HNAS+GVSSACPFNG SSFFRNVVWLIKKNDVYRTLKR
Sbjct  119  HLMSSTNHLERIQIFPRNSWSSHNASSGVSSACPFNGRSSFFRNVVWLIKKNDVYRTLKR  178

Query  119  NYTNTNVEDLLIIWGVHHPNDATEQTKLYQNLNTYVSVGTSTLNQRSTPTIATRPQVNGQ  178
            NYTNTNVEDLLIIWGVHHPNDATEQTKLYQNLNTYVSVGTSTLNQRS PTIATRPQVNGQ
Sbjct  179  NYTNTNVEDLLIIWGVHHPNDATEQTKLYQNLNTYVSVGTSTLNQRSIPTIATRPQVNGQ  238

Query  179  RGRMEFFWTILKSNDSIFFESTGNFIAPEYAYKIVKKGKSTVMK  222
            RGRMEFFWTILKSNDSIFFESTGNFIAPEYAYKIVKKG S VMK
Sbjct  239  RGRMEFFWTILKSNDSIFFESTGNFIAPEYAYKIVKKGNSAVMK  282
```

Modelling and Structure comparison for Avian HA-RBD Subunit 1

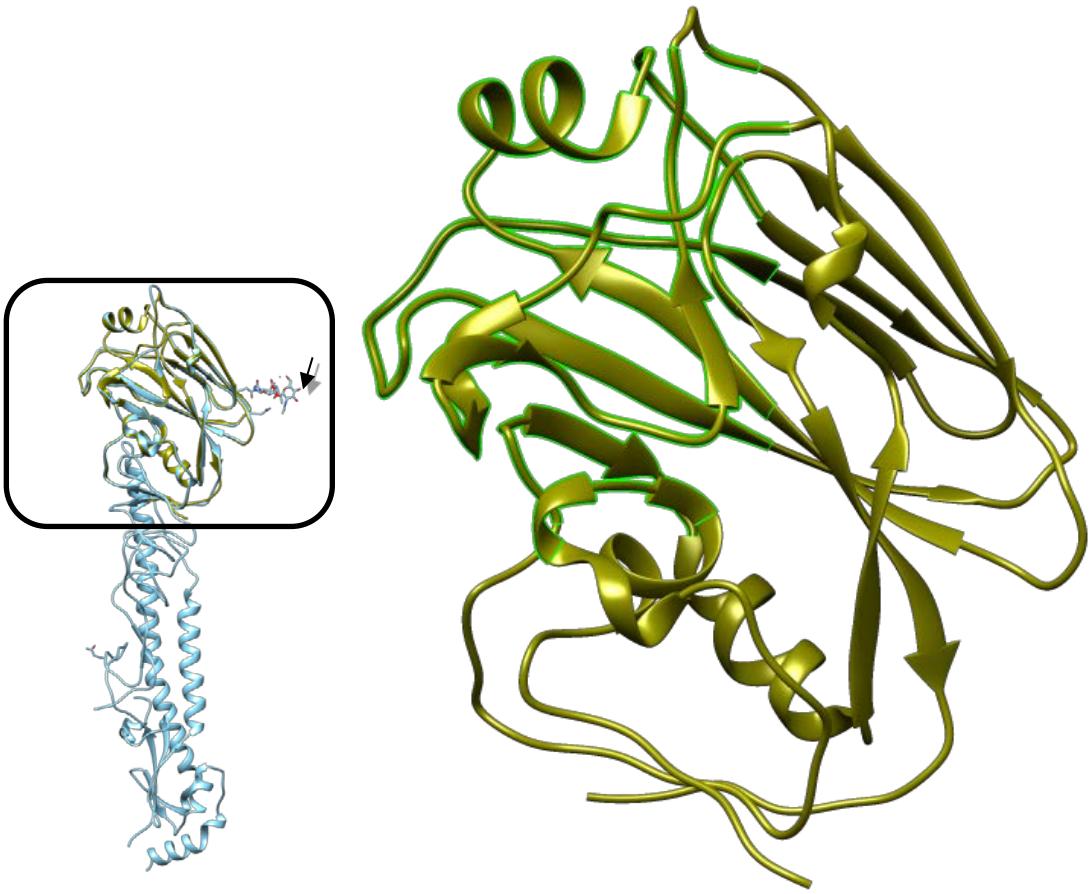
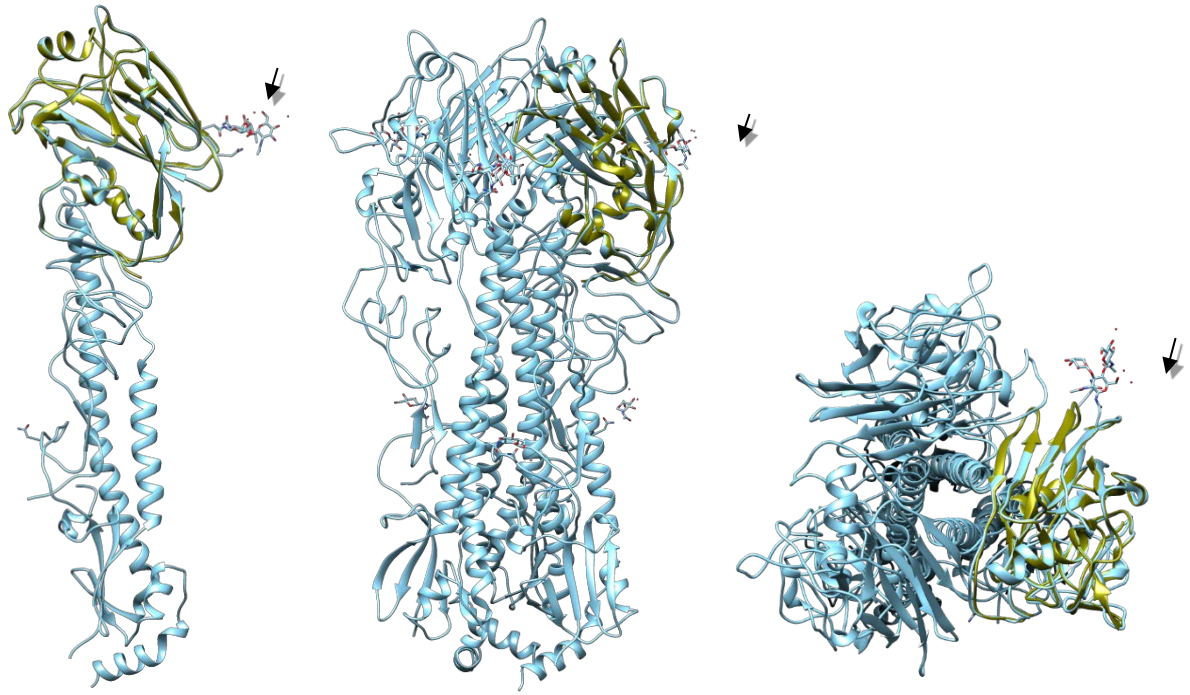


Figure 13. Structural model of the Avian HA RBD subunit 1. The PDB file of globular head comprising the Avian HA-RBD subunit 1 (in olive green) was generated by the Phyre2 server (5) and superimposed over the determined crystal structure map of the Hemagglutinin HA5 (in cyan) of the virus A/duck/Egypt/10185SS/2010 (H5N1) (PDB file 5E2Y) for comparison and illustration of the region selected for antigen expression in the present study (black box). The arrow indicates the N- acetyl-D-glucosamine molecule, interacting with one of the receptor binding sites on the HA5 molecule. The graphs were generated using Chimera UCSF software, and the crystal structure PDB file was retrieved from the RCSB Protein Data Bank website.



A B C

Figure 14. Structural model of the Avian HA RBD subunit 1. The PDB file of the molecule generated for the Avian HA RBD Subunit 1 (in olive) was superimposed over the crystal structure determined map of the Hemagglutinin HA5 (in cyan) of the virus A/duck/Egypt/10185SS/2010 (H5N1) (PDB file 5E2Y) for comparison and illustration of the region selected for antigen expression in the present study (9). In A, the Avian RBD1 was superimposed on the whole HA monomeric molecule; B, the Avian RBD1 superimposed over one monomer of the hemotrimeric HA5 form; C, Upper view of C. Arrows indicated the N-Acetyl-D-glucosamine molecule, interacting with the receptor binding site of HA5 molecule. The PDB structure file of the Avian HA RBD subunit 1 was generated using the Phyre2 server, analysis and superposition of graphs generated using Chimera UCSF software. The crystal structure PDB file was retrieved from the RCSB Protein Data Bank website.

Table 6. Protein Sequence alignment of the Avian HA RBD Subunit 1 compared to the X-Ray diffraction data for HA1 domain of A/duck/Egypt/10185SS/2010 (H5N1) virus. (*) asterisk indicated fully conserved residue; (:), semicolon, conservation among groups of residues with strongly similar propiertes; (.), period indicates conservation residues conservation among weakly similar propierties. The aligment was generated using ClustalW program from the Chimera UCSF software.

```
CLUSTAL W ALN saved from UCSF Chimera/MultAlignViewer

5e2y,_chain_A/7-339               ADPGDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCNLDG 50
Avian_RBD1_complete.pdb/5-224     ..................................................


5e2y,_chain_A/7-339               VKPLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKINPANDLCYPGNFN 100
Avian_RBD1_complete.pdb/5-224     KGGFILKDCSVAGWLLGNPMCDEFLTAPEWSYIVEKDNPSNDLCYPGNFD 50
                                     :**:******************..********* **:**********:

5e2y,_chain_A/7-339               DYEELKHLLSRINHFEKIQITPKNSWSDHEAS.GVSSACPYQGRSSFFRN 149
Avian_RBD1_complete.pdb/5-224     NYEELKHLMSSTNHMERIQIFPRNSWTNHNASAGVSSACPFNGGSSFFRN 100
                                  :*******:*  **:*:*** *:***::*:** *******::* ******

5e2y,_chain_A/7-339               VVWLTKKDNAYPTIKRSYNNTNQEDLLVLWGIHHPNDATEQTRLYQNPTT 199
Avian_RBD1_complete.pdb/5-224     VVWLIKKNDVYRTLKRNYTNTNVEDLLIIWGVHHPNDATEQTKLYQNLNT 150
                                  **** **::.* *:**.*.*** ****::**:**********:**** .*

5e2y,_chain_A/7-339               YISVGTSTLNQKLVPKIATRSKVKGLSGRMEFFWTILKSNDAINFESNGN 249
Avian_RBD1_complete.pdb/5-224     YVSVGTSTLNQRSTPTIATRPQVNGQRGRMEFFWTILKSNDSIFFESTGN 200
                                  *:*********: .*.****.:*:*  **************:* ***.**

5e2y,_chain_A/7-339               FIAPENAYKIVKKGDSTIMKSELEYGDCNTKCQTPIGAINSSMPFHNIHP 299
Avian_RBD1_complete.pdb/5-224     FIAPEYAYKIVKKGKSTVMK.............................. 220
                                  ***** ********.**:**

5e2y,_chain_A/7-339               LTIGECPKYVKSNRLVLATGLRNSPQGERRRKKR 333
Avian_RBD1_complete.pdb/5-224     ................................
```

Table 7. The HA-RBD Subunit 2 Sequence was derived from a field virus isolation from chicken farms located in Mexico. The sequences provided by IASA company were used in Blast searches of the NCBI database to compare and identify the origin of the Hemagglutinin RBD gene product used in this study. The HA RBD subunit 2 has a high protein homology to the H5-like hemaglutinin protein. The best score was obtained for the Influenza A virus -A/chicken/Hidalgo/1336-06/2006(H5N2)- isolate present in the database.

Hemaglutinin [Influenza A virus (A/chicken/Hidalgo/1336-06/2006(H5N2))]

Sequence ID: gi|682021806|gb|AIN25320.1|Length: 564 Number of Matches: 1

Score	Expect	Method	Identities	Positives	Gaps
437 bits(1123)	5e-150	Compositional matrix adjust.	205/218(94%)	212/218(97%)	0/218(0%)

```
Query  1    PLILKDCSVAGWLLGNPMCDEFLNVQEWSYIVEKDNPSNDLCYPGNFDNYEELKHLMSST  60
            PLILKDCSVAGWLLGNPMCDEFL+V EWSYIVEKDNPSNDLCYPGNFDNYEELKHLMSST
Sbjct  65   PLILKDCSVAGWLLGNPMCDEFLSVPEWSYIVEKDNPSNDLCYPGNFDNYEELKHLMSST  124

Query  61   NHLERIQIFPRNSWPNHDSSSGVSSACPFNGESSFFRNVVWLIKKNDVYRTLKRNYTNTN  120
            NHLERIQIFPRNSW +H++SSGVSSACPFNG SSFFRNVVWLIKKNDVYRTLKRNYTNTN
Sbjct  125  NHLERIQIFPRNSWSSHNASSGVSSACPFNGRSSFFRNVVWLIKKNDVYRTLKRNYTNTN  184

Query  121  VEDLLIIWGVHHPNDATVQTKLYQNLNTYVSVGTSTLNQRSIPTIATRPQVNGQRGRVEF  180
            VEDLLIIWGVHHPNDAT QTKLYQNLNTYVSVGTSTLNQRSIPTIATRPQVNGQRGR+EF
Sbjct  185  VEDLLIIWGVHHPNDATEQTKLYQNLNTYVSVGTSTLNQRSIPTIATRPQVNGQRGRMEF  244

Query  181  FWTILESNDSIFFESTGNFIAPEYAYKVTKKGKSAVMK  218
            FWTIL+SNDSIFFESTGNFIAPEYAYK+ KKG SAVMK
Sbjct  245  FWTILKSNDSIFFESTGNFIAPEYAYKIVKKGNSAVMK  282
```

Modelling and Structure comparison for Avian HA-RBD Subunit 2

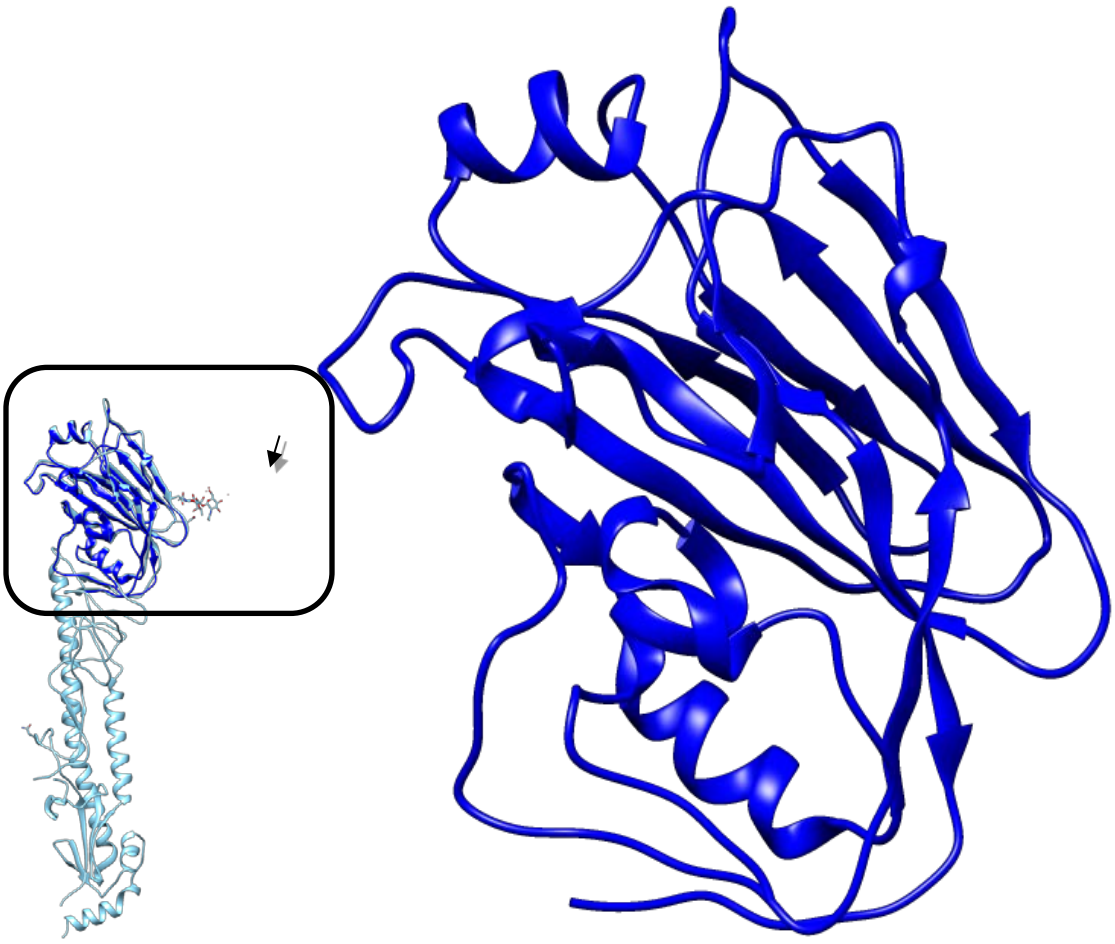


Figure 15. Structural model of the Avian HA-RBD subunit 2. The PDB file of the globular head comprising the Avian HA-RBD subunit 2 (in blue) and generated by the Phyre2 server (5), was superimposed over the determined crystal structure map of the Hemagglutinin HA5 (in cyan) of the virus A/duck/Egypt/10185SS/2010 (H5N1) (PDB file 5E2Y) for comparison and illustration of the region selected for antigen expression in the present study (black box) (9). Arrow indicates the N-acetyl-D-glucosamine molecule, interacting with one of the receptor binding sites on the HA5 molecule. The graphs were generated using Chimera UCSF software, crystal structure PDB file retrieved from the RCSB Protein Data Bank website.

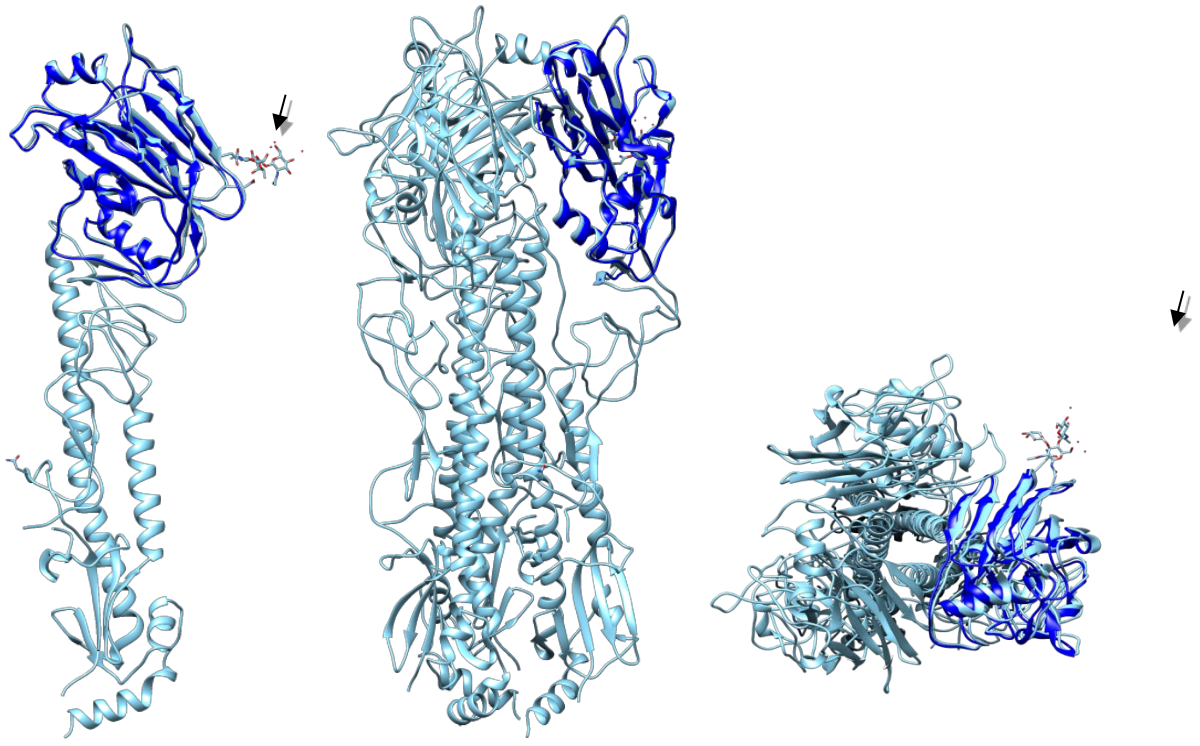


Figure 16. Structural model of the Avian HA RBD subunit 2. The generated PDB file of the Avian HA RBD Subunit 2 molecule (in blue) was superimposed over the crystal structure-determined map of the Hemagglutinin HA5 (in cyan) of the virus A/duck/Egypt/10185SS/2010 (H5N1) (PDB file 5E2Y) for comparison and illustration of the region selected for antigen expression in the present study. In A, Avian RBD2 superimposed on whole HA H5 monomeric molecule; B, Avian RBD2 superimposed over one monomer of the hemotrimeric HA H5 form; C, Upper view of C. Arrows indicated the NAG (N- acetyl-D-glucosamine) molecule, interacting with the receptor binding site of HA H5 molecule. PDB of the Avian HA RBD subunit 2 was generated by Phyre2 server, analysis and superposition graphs was generated using Chimera UCSF software, crystal structure PDB file was retrieved from the RCSB Protein Data Bank website.

Table 8. Protein Sequence alignment of the Avian HA-RBD Subunit 2 compared to the X-Ray diffraction HA1 domain of A/duck/Egypt/10185SS/2010 (H5N1) (PDB 5e2y) virus. (*) asterisk indicated fully conserved residue; (:), semicolon, conservation among groups of residues with strongly similar propiertes; (.), period indicates conservation residues conservation among weakly similar propierties. The aligment was generated using ClustalW program from the Chimera UCSF software.

```
CLUSTAL W ALN saved from UCSF Chimera/MultAlignViewer

5e2y,_chain_A/7-339               ADPGDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCNLDG 50
Avian_RBD2_complete.pdb/1-218     ..................................................


5e2y,_chain_A/7-339               VKPLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKINPANDLCYPGNFN 100
Avian_RBD2_complete.pdb/1-218     ..PLILKDCSVAGWLLGNPMCDEFLNVQEWSYIVEKDNPSNDLCYPGNFD 48
                                    ****:******************** ******** **:*********:

5e2y,_chain_A/7-339               DYEELKHLLSRINHFEKIQITPKNSWSDHEAS.GVSSACPYQGRSSFFRN 149
Avian_RBD2_complete.pdb/1-218     NYEELKHLMSSTNHLERIQIFPRNSWPNHDSSSGVSSACPFNGESSFFRN 98
                                  :*******:*  **:*:*** *:***.:*::* *******::*.******

5e2y,_chain_A/7-339               VVWLTKKDNAYPTIKRSYNNTNQEDLLVLWGIHHPNDATEQTRLYQNPTT 199
Avian_RBD2_complete.pdb/1-218     VVWLIKKNDVYRTLKRNYTNTNVEDLLIIWGVHHPNDATVQTKLYQNLNT 148
                                  **** **::.* *:**.*.*** ****::**:******* **:**** .*

5e2y,_chain_A/7-339               YISVGTSTLNQKLVPKIATRSKVKGLSGRMEFFWTILKSNDAINFESNGN 249
Avian_RBD2_complete.pdb/1-218     YVSVGTSTLNQRSIPTIATRPQVNGQRGRVEFFWTILESNDSIFFESTGN 198
                                  *:*********: :*.****.:*:*  **:*******:***:* ***.**

5e2y,_chain_A/7-339               FIAPENAYKIVKKGDSTIMKSELEYGDCNTKCQTPIGAINSSMPFHNIHP 299
Avian_RBD2_complete.pdb/1-218     FIAPEYAYKVTKKGKSAVMK.............................. 218
                                  ***** ***:.***.*::**

5e2y,_chain_A/7-339               LTIGECPKYVKSNRLVLATGLRNSPQGERRRKKR 333
Avian_RBD2_complete.pdb/1-218     ................................
```

.

Table 9. The Avian HA-RBD Subunit 3 Sequence was derived from field virus isolates obtained from chicken farms located in Mexico. The sequences provided by IASA company were used in Blast searches of the NCBI database to compare and identify the origin of the Hemagglutinin RBD gene product used in this study (Query 1). HA-RBD subunit 3 has a high protein homology to H7-like hemaglutinin protein (Sbjct) in the present case, the best score was obtained for the Influenza A virus (A/Turkey/Ontario/18/2000(H7) virus isolate.

Hemagglutinin, partial [Influenza A virus (A/Turkey/Ontario/18/2000(H7))]

Sequence ID: gi|20335019|gb|AAM19229.1|AF497552_1Length: 337Number of Matches: 1

Score	Expect	Method	Identities	Positives	Gaps
424 bits(1091)	2e-148	Compositional matrix adjust.	202/213(95%)	206/213(96%)	0/213(0%)

```
Query  1    PTDLGQCGLLGTLIGPPQCDQFLKFDADLIIERREGTDVCYPGKFTNEESLRQILRRSGG  60
            PTDLGQCGLLGTLIGPPQCDQFL+FDADLIIERREGTDVCYPGKFTNEESLRQILR SGG
Sbjct  53   PTDLGQCGLLGTLIGPPQCDQFLEFDADLIIERREGTDVCYPGKFTNEESLRQILRGSGG  112

Query  61   IDKESMGFTYSGIRTNGTTSACRRSNPSFYAEMKWLLSNSDNATFPQMTKSYRNPRNKPA  120
            IDKESMGFTYSGIRTNG TSACRRS  SFYAEMKWLLSNSDNA FPQMTKSYRNPRNKPA
Sbjct  113  IDKESMGFTYSGIRTNGATSACRRSGSSFYAEMKWLLSNSDNAAFPQMTKSYRNPRNKPA  172

Query  121  LITWGVHHSGSATEQTKLYGSGDKLITVGSSKYLQSFTPSPGARPQVNGQSGRIDFHWLL  180
            LI WGVHHSGSATEQTKLYGSG+KLITVGSSKY QSFTPSPGARPQVNGQSGRIDFHWLL
Sbjct  173  LIIWGVHHSGSATEQTKLYGSGNKLITVGSSKYQQSFTPSPGARPQVNGQSGRIDFHWLL  232

Query  181  LDPNDTVTFTFNGAFIAPDRASFFRGESIGIQS  213
            LDPNDTVTFTFNGAFIAPDRASFFRGES+G+QS
Sbjct  233  LDPNDTVTFTFNGAFIAPDRASFFRGESLGVQS  265
```

Modelling and Structure comparison for Avian HA-RBD Subunit 3

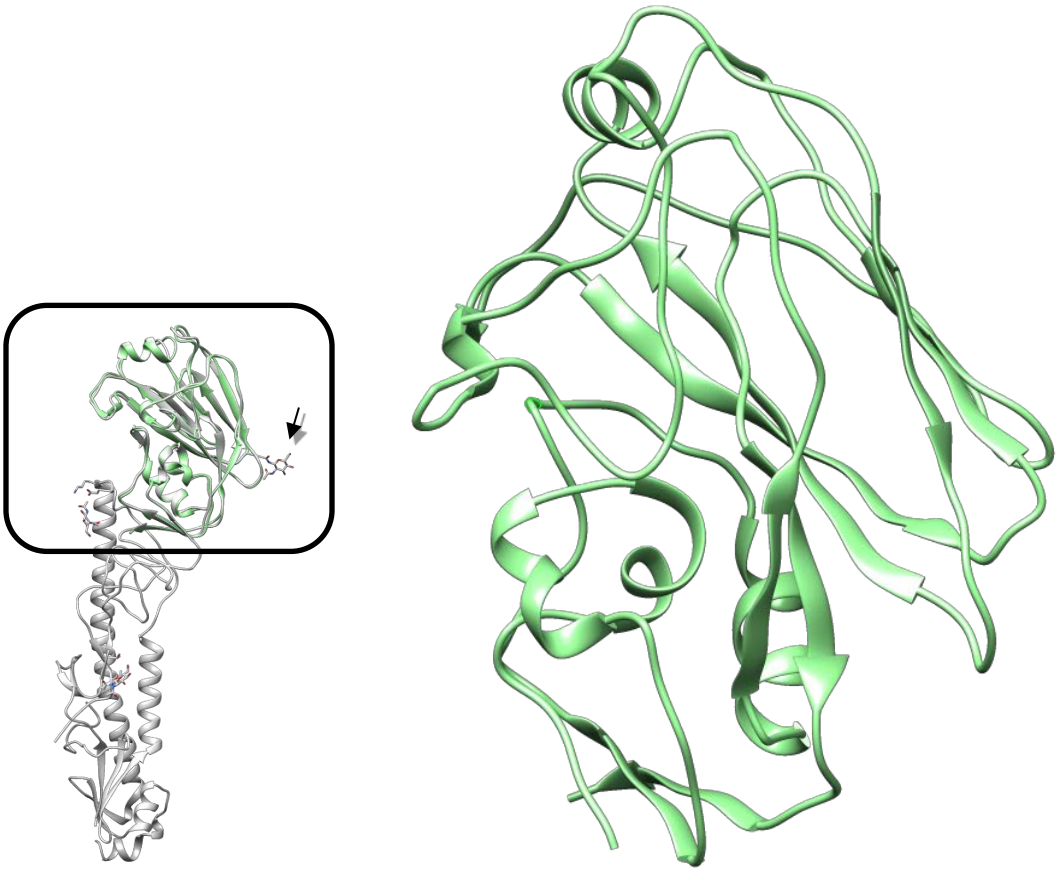


Figure 17. Structural model of the Avian HA-RBD subunit 3. The PDB file of globular head comprising the Avian HA-RBD Subunit 3 (in green) and generated by the Phyre2 server (5), superimposed over the determined crystal structure map of the Hemagglutinin H7 (in cyan,) of the virus A/Anhui/1/2013 (PDB file 4R8W) (10) for comparison and illustration of the region selected for antigen expression in the present study (black box). Arrow indicates the NAG (N-acetyl-D-glucosamine) molecule, interacting with one receptor binding site of HA5 molecule. The graphs were generated using Chimera UCSF software, crystal structure PDB file retrieved from the RCSB Protein Data Bank website.

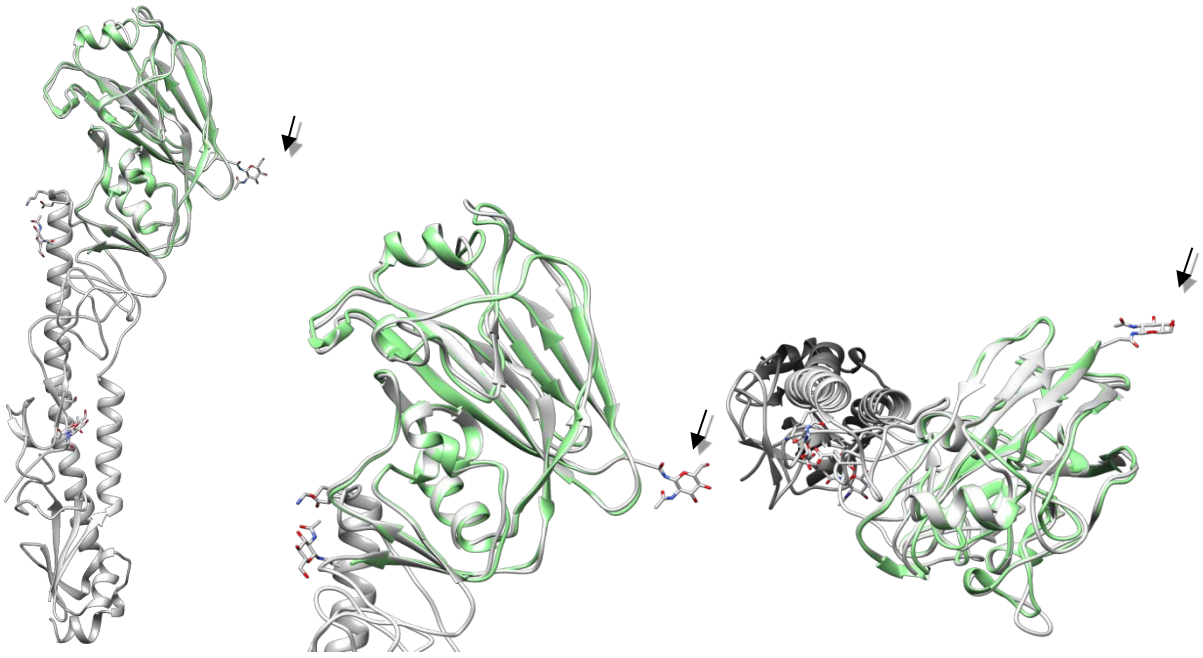


Figure 18. Structural model of the Avian HA-RBD subunit 3. The generated PDB file of the molecule Avian HA-RBD Subunit 3 (in blue) superimposed over the crystal structure determined map of the Hemagglutinin H7 (in cyan) of the virus A/Anhui/1/2013 (PDB file 4R8W) (10) for comparison and illustration of the region selected for antigen expression in the present study. In A, Avian RBD2 superimposed on whole HA H5 monomeric molecule; B, Avian RBD2 superimposed over one monomer of the hemotrimeric HA H5 form; C, Upper view of C. Arrows indicates the NAG (N- acetyl-D-glucosamine) molecule, interacting with the receptor binding site of HA H5 molecule. PDB of the Avian HA RBD subunit 2 was generated by Phyre2 server (5) , analysis and superposition graphs was generated using Chimera UCSF software, crystal structure PDB file was retrieved from the RCSB Protein Data Bank website.

Table 10. Protein Sequence alignment of the Avian HA-RBD Subunit 3 compared to the X-Ray diffraction HA1 domain of H7 Hemagglutinin of A/Anhui/1/2013 virus. (*) asterisk indicated fully conserved residue; (:), semicolon, conservation among groups of residues with strongly similar propiertes; (.), period indicates conservation residues conservation among weakly similar properties. The alignment was generated using ClustalW program from the Chimera UCSF software.

```
CLUSTAL W ALN saved from UCSF Chimera/MultAlignViewer

4R8W,_chain_A/1-321               DKICLGHHAVSNGTKVNTLTERGVEVVNATETVERTNIPRICSKGKRTVD 50
Avian_RBD3_complete.pdb/1-213     ...............................................PTD 3
                                                                                 ..*

4R8W,_chain_A/1-321               LGQCGLLGTITGPPQCDQFLEFSADLIIERREGSDVCYPGKFVNEEALRQ 100
Avian_RBD3_complete.pdb/1-213     LGQCGLLGTLIGPPQCDQFLKFDADLIIERREGTDVCYPGKFTNEESLRQ 53
                                  *********: *********:*.**********:********.***:***

4R8W,_chain_A/1-321               ILRESGGIDKEAMGFTYSGIRTNGATSACRRSGSSFYAEMKWLLSNTDNA 150
Avian_RBD3_complete.pdb/1-213     ILRRSGGIDKESMGFTYSGIRTNGTTSACRRSNPSFYAEMKWLLSNSDNA 103
                                  ***.*******:************:*******..************:***

4R8W,_chain_A/1-321               AFPQMTKSYKNTRKSPALIVWGIHHSVSTAEQTKLYGSGNKLVTVGSSNY 200
Avian_RBD3_complete.pdb/1-213     TFPQMTKSYRNPRNKPALITWGVHHSGSATEQTKLYGSGDKLITVGSSKY 153
                                  :********:*.*:.****.**:*** *::*********:**:*****:*

4R8W,_chain_A/1-321               QQSFVPSPGARPQVNGLSGRIDFHWLMLNPNDTVTFSFNGAFIAPDRASF 250
Avian_RBD3_complete.pdb/1-213     LQSFTPSPGARPQVNGQSGRIDFHWLLLDPNDTVTFTFNGAFIAPDRASF 203
                                  ***.********** *********:*:*******:*************

4R8W,_chain_A/1-321               LRGKSMGIQSGVQVDANCEGDCYHSGGTIISNLPFQNIDSRAVGKCPRYV 300
Avian_RBD3_complete.pdb/1-213     FRGESIGIQS..................................... 213
                                  :**:*:****

4R8W,_chain_A/1-321               KQRSLLLATGMKNVPEIPKGR 321
Avian_RBD3_complete.pdb/1-213     .....................
```

Table 11. Protein Sequence alignment of the Avian HA-RBD Subunits used to construct the Avian multimeric gene versions. The RBD1, RBD2, RBD3 were aligned for comparison to IV A/California/07/2009(H1N1), which was used as a template to determine the border sequences that determine the lectin domain and was based on the conserved residues (in blue) shared among all the sequences as the backbone scaffold; their predicted globular folding were in Figures 13-17. (*) asterisk indicated fully conserved residue; (:), semicolon, conservation among groups of residues with strongly similar properties; (.), period indicates conservation residues conservation among weakly similar properties; (highlighted in grey) sequence corresponding to the fusion peptide present in HA2 domain. The alignment was generated using ClustalW program from the Chimera UCSF software.

```
CLUSTAL W ALN saved from UCSF Chimera/MultAlignViewer

A/California/07/2009(H1N1)      MKAILVVLLYTFATANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLL 50
Avian_RBD1_(H5-like)            --------------------------------------------------
Avian_RBD2_(H5-like)            --------------------------------------------------
Avian_RBD3_(H7-like)            --------------------------------------------------


A/California/07/2009(H1N1)      EDKHNGKLCKLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETP 100
Avian_RBD1_(H5-like)            ----------GGFILKDCSVAGWLLGNPMCDEFLTAPEWSYIVEKD 37
Avian_RBD2_(H5-like)            ----------SPLILKDCSVAGWLLGNPMCDEFLNVQEWSYIVEKD 36
Avian_RBD3_(H7-like)            ----------SPTDLGQCGLLGTLIGPPQCDQFLKFDA-DLIIER- 34
                                          .   * .*.: * ::* * *:.: .    . *:*

A/California/07/2009(H1N1)      SSDNGTCYPGDFIDYEELREQLSSVSSFERFEI-FPKTSSWPNHDSNKGV 149
Avian_RBD1_(H5-like)            NPSNDLCYPGNFDNYEELKHLMSSTNHMERIQI-FPR-NSWTNHNASAGV 85
Avian_RBD2_(H5-like)            NPSNDLCYPGNFDNYEELKHLMSSTNHLERIQI-FPR-NSWPNHDSSSGV 84
Avian_RBD3_(H7-like)            REGTDVCYPGKFTNEESLRQILRRSGGIDKESMGFTYSGIRTN-----GT 79
                                  ... ****.* : *.*:. :   . ::: .: *.  .  .*     *.

A/California/07/2009(H1N1)      TAACPHAGAKSFYKNLIWLVKKGN--SYPKLSKSYINDKGKEVLVLWGIH 197
Avian_RBD1_(H5-like)            SSACPFNGGSSFFRNVVWLIKKND--VYRTLKRNYTNTNVEDLLIIWGVH 133
Avian_RBD2_(H5-like)            SSACPFNGESSFFRNVVWLIKKND--VYRTLKRNYTNTNVEDLLIIWGVH 132
Avian_RBD3_(H7-like)            TSAC-RRSNPSFYAEMKWLLSNSDNATFPQMTKSYRNPRNKPALITWGVH 128
                                ::**   .  **: :: **:.:.:   :  :.:.* * . :  *: **:*

A/California/07/2009(H1N1)      HPSTSADQQSLYQNADAYVFVGSSRYSKKFKPEIAIRPKVRDQEGRMNYY 247
Avian_RBD1_(H5-like)            HPNDATEQTKLYQNLNTYVSVGTSTLNQRSTPTIATRPQVNGQRGRMEFF 183
Avian_RBD2_(H5-like)            HPNDATVQTKLYQNLNTYVSVGTSTLNQRSIPTIATRPQVNGQRGRVEFF 182
Avian_RBD3_(H7-like)            HSGSATEQTKLYGSGDKLITVGSSKYLQSFTPSPGARPQVNGQSGRIDFH 178
                                *.. :: * .** . :  : **:*   :   *  . **:*..* **:::.

A/California/07/2009(H1N1)      WTLVEPGDKITFEATGNLVVPRYAFAMERNAGSGIIISDTPVHDCNTTCQ 297
Avian_RBD1_(H5-like)            WTILKSNDSIFFESTGNFIAPEYAYKIVKKGKSTVMKS------------ 220
Avian_RBD2_(H5-like)            WTILESNDSIFFESTGNFIAPEYAYKVTKKGKSAVMKS------------ 219
Avian_RBD3_(H7-like)            WLLLDPNDTVTFTFNGAFIAPDRASFF--RGESIGIQS------------ 214
                                * ::...*.: *  .* ::.*  *  .  .. *  : *

A/California/07/2009(H1N1)      TPKGAINTSLPFQNIHPITIGKCPKYVKSTKLRLATGLRNIPSIQSRGLF 347
Avian_RBD1_(H5-like)            --------------------------------------------------
Avian_RBD2_(H5-like)            --------------------------------------------------
Avian_RBD3_(H7-like)            --------------------------------------------------


A/California/07/2009(H1N1)      GAIAGFIEGGWTGMVDGWYGYHHQNEQGSGYAADLKSTQNAIDEITNKVN 397
Avian_RBD1_(H5-like)            --------------------------------------------------
Avian_RBD2_(H5-like)            --------------------------------------------------
Avian_RBD3_(H7-like)            --------------------------------------------------


A/California/07/2009(H1N1)      SVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWTYNAELLVLLE 447
Avian_RBD1_(H5-like)            --------------------------------------------------
Avian_RBD2_(H5-like)            --------------------------------------------------
Avian_RBD3_(H7-like)            --------------------------------------------------
```

```
A/California/07/2009(H1N1)      NERTLDYHDSNVKNLYEKVRSQLKNNAKEIGNGCFEFYHKCDNTCMESVK 497
Avian_RBD1_(H5-like)            --------------------------------------------------
Avian_RBD2_(H5-like)            --------------------------------------------------
Avian_RBD3_(H7-like)            --------------------------------------------------


A/California/07/2009(H1N1)      NGTYDYPKYSEEAKLNREEIDGVKLESTRIYQILAIYSTVASSLVLVVSL 547
Avian_RBD1_(H5-like)            --------------------------------------------------
Avian_RBD2_(H5-like)            --------------------------------------------------
Avian_RBD3_(H7-like)            --------------------------------------------------



A/California/07/2009(H1N1)      GAISFWMCSNGSLQCRICI 566
Avian_RBD1_(H5-like)            ------------------
Avian_RBD2_(H5-like)            ------------------
Avian_RBD3_(H7-like)            ------------------
```

3.1.3. HIV-TAT cell penetrating peptide sequence

The correct synthesis of the nucleotide sequence coding for the Tat-cell penetrating peptide *RKKRRQRRR* was confirmed in the sequencing results of the Avian HA-RBD V3.7 version (**Table 12**) retrieved by GeneScript Inc.

3.1.4. 5-mer Immuno-stimulatory peptide sequence

The correct synthesis of the nucleotide sequence coding for the 5-mer immune-stimulatory peptide *KWCEC* was confirmed in the sequencing results of the Avian HA-RBD V3.6 (**Table 11**) and Avian HA-RBD V3.7 (**Table 12**) versions retrieved by GeneScript Inc.

3.2. Avian RBD designs, synthetic gene sequences

The codon optimized DNA sequence of the HA RBD CV3.4 was cloned into the yeast *Pinchia pastoris* strain X-33.

Sequencing results confirmed the correct codon-optimized gene synthesis and their specific features for the Avian HA-RBD V3.4 multimeric gene, 2026 nucleotide base pairs (**Table 10**, **Table 14**); for Avian HA-RBD V3.6 multimeric gene, 2055 nucleotide base pairs (**Table 11**, **Table 14**); for Avian HA-RBD V3.7 multimeric gene, 2088 nucleotide base pairs (**Table 12**, **Table 14**); and finally for the Human HA-RBD V3.3 multimeric gene, 2145 nucleotide base pairs (**Table 13**, **Table 14**). All gene versions were custom cloned in the extracellular vector pPicZalphaA flanked at the 5' site by the restriction sequence XhoI and at the 3' by the restriction site XbaI for all versions.

Table 12. The Avian HA-RBD V3.4 sequence was synthetized and codon optimized for *Pichia pastoris* expression using the OptiGene software available from GeneScript Inc. Features of the synthetized sequence include: an XhoI, 5' restriction site; an XbaI, 3' restriction site; a pSTOP, premature stop codon; an RBD1, Receptor binding domain derived from H5; an RBD2, the Receptor binding domain derived from H5 an; RBD3, and the Receptor binding domain DNA sequence derived from H7 hemagglutinin. Sequence analysis and its theoretical translation to protein were performed using the software DNAMAN V9.0 (Lynnon). The list of features was obtained through SnapGene software (SnapGene).

Translation of HA **RBD V(3.4)**(1-2046)
Universal code H

Total amino acid number: 680, MW=75697
Max ORF starts at AA pos 1(may be DNA pos 1) for 679 AA(2037 bases), MW=75610

```
                 10        20        30        40        50        60
1         CTCGAGAAGAGAAGTGTTAAGGGAGGTTTTATTTTGAAAGACTGTTCTGTCGCTGGTTGG
1          L  E  K  R  S  V  K  G  G  F  I  L  K  D  C  S  V  A  G  W

                 70        80        90       100       110       120
61        TTGCTTGGAAATCCTATGTGTGATGAGTTTTTGACTGCTCCAGAATGGTCCTACATTGTT
21         L  L  G  N  P  M  C  D  E  F  L  T  A  P  E  W  S  Y  I  V

                130       140       150       160       170       180
121       GAGAAGGATAACCCATCAAATGACTTGTGTTACCCTGGTAACTTCGATAACTACGAAGAG
41         E  K  D  N  P  S  N  D  L  C  Y  P  G  N  F  D  N  Y  E  E

                190       200       210       220       230       240
181       CTTAAGCATTTGATGTCTTCCACCAACCACATGGAAAGAATTCAAATCTTCCCAAGAAAT
61         L  K  H  L  M  S  S  T  N  H  M  E  R  I  Q  I  F  P  R  N

                250       260       270       280       290       300
241       TCCTGGACTAACCATAATGCTTCAGCCGGTGTTTCAAGTGCTTGCCCTTTTAACGGTGGA
81         S  W  T  N  H  N  A  S  A  G  V  S  S  A  C  P  F  N  G  G

                310       320       330       340       350       360
301       TCTTCCTTTTTCAGAAATGTTGTCTGGCTTATTAAGAAAAACGATGTTTACAGAACTTTG
101        S  S  F  F  R  N  V  V  W  L  I  K  K  N  D  V  Y  R  T  L

                370       380       390       400       410       420
361       AAGAGAAACTACACCAACACTAATGTTGAGGATTTGCTTATTATCTGGGGTGTCCATCAC
121        K  R  N  Y  T  N  T  N  V  E  D  L  L  I  I  W  G  V  H  H

                430       440       450       460       470       480
421       CCAAATGACGCTACTGAACAAACAAAGCTTTACCAGAACTTGAACACATACGTTAGTGTC
141        P  N  D  A  T  E  Q  T  K  L  Y  Q  N  L  N  T  Y  V  S  V

                490       500       510       520       530       540
481       GGTACATCTACCTTGAACCAAAGATCAACTCCAACAATTGCCACTAGACCTCAAGTTAAT
161        G  T  S  T  L  N  Q  R  S  T  P  T  I  A  T  R  P  Q  V  N

                550       560       570       580       590       600
541       GGTCAGAGAGGAAGAATGGAATTTTTCTGGACTATCTTGAAGTCCAACGATTCAATCTTT
181        G  Q  R  G  R  M  E  F  F  W  T  I  L  K  S  N  D  S  I  F
```

```
                610       620       630       640       650       660
601       TTCGAGTCCACTGGTAATTTTATTGCCCCTGAATACGCATACAAGATCGTTAAGAAAGGA
201        F  E  S  T  G  N  F  I  A  P  E  Y  A  Y  K  I  V  K  K  G

                670       680       690       700       710       720
661       AAGTCCACTGTCATGAAATCTGGTGGAGGTGGAAGTGGTGGAGGTGGATCTCCACTTATT
221        K  S  T  V  M  K  S  G  G  G  G  S  G  G  G  G  S  P  L  I

                730       740       750       760       770       780
721       TTGAAGGATTGTTCTGTTGCTGGATGGTTGCTTGGTAATCCTATGTGCGACGAATTCTTG
241        L  K  D  C  S  V  A  G  W  L  L  G  N  P  M  C  D  E  F  L

                790       800       810       820       830       840
781       AACGTTCAAGAGTGGTCTTACATCGTCGAAAAGGATAACCCATCCAACGACTTGTGCTAT
261        N  V  Q  E  W  S  Y  I  V  E  K  D  N  P  S  N  D  L  C  Y

                850       860       870       880       890       900
841       CCAGGAAATTTTGACAATTATGAGGAGCTTAAACATTTGATGTCAAGTACTAACCACTTG
281        P  G  N  F  D  N  Y  E  E  L  K  H  L  M  S  S  T  N  H  L

                910       920       930       940       950       960
901       GAGAGAATTCAGATCTTCCCAAGAAACTCTTGGCCTAATCACGATTCTTCCTCAGGTGTT
301        E  R  I  Q  I  F  P  R  N  S  W  P  N  H  D  S  S  S  G  V

                970       980       990      1000      1010      1020
961       AGTTCTGCTTGTCCTTTTAACGGAGAATCCTCATTTTTCAGAAACGTTGTCTGGTTGATT
321        S  S  A  C  P  F  N  G  E  S  S  F  F  R  N  V  V  W  L  I

               1030      1040      1050      1060      1070      1080
1021      AAGAAAAATGACGTCTATAGAACCCTTAAGAGAAACTATACCAACACTAATGTCGAAGAT
341        K  K  N  D  V  Y  R  T  L  K  R  N  Y  T  N  T  N  V  E  D

               1090      1100      1110      1120      1130      1140
1081      TTGTTGATCATCTGGGGTGTTCATCACCCTAATGACGCAACCGTCCAAACTAAATTGTAC
361        L  L  I  I  W  G  V  H  H  P  N  D  A  T  V  Q  T  K  L  Y

               1150      1160      1170      1180      1190      1200
1141      CAGAACTTGAACACTTACGTTTCCGTCGGTACATCAACCTTGAACCAAAGATCTATTCCA
381        Q  N  L  N  T  Y  V  S  V  G  T  S  T  L  N  Q  R  S  I  P

               1210      1220      1230      1240      1250      1260
1201      ACAATCGCTACCAGACCACAGGTCAACGGTCAGAGAGGAAGAGTCGAGTTTTTCTGGACC
401        T  I  A  T  R  P  Q  V  N  G  Q  R  G  R  V  E  F  F  W  T

               1270      1280      1290      1300      1310      1320
1261      ATTTTGGAAAGTAACGATTCTATCTTTTTCGAGTCTACTGGTAACTTCATTGCTCCAGAA
421        I  L  E  S  N  D  S  I  F  F  E  S  T  G  N  F  I  A  P  E

               1330      1340      1350      1360      1370      1380
1321      TACGCCTATAAGGTTACAAAGAAAGGAAAGTCCGCTGTCATGAAGTCTGGAGGTGGAGGT
441        Y  A  Y  K  V  T  K  K  G  K  S  A  V  M  K  S  G  G  G  G

               1390      1400      1410      1420      1430      1440
1381      TCTGGAGGAGGTGGATCTCCTACCGATCTTGGTCAATGTGGATTGCTTGGTACTTTGATT
461        S  G  G  G  G  S  P  T  D  L  G  Q  C  G  L  L  G  T  L  I
               1450      1460      1470      1480      1490      1500
```

```
1441      GGTCCACCTCAATGCGACCAGTTTCTTAAATTCGATGCCGACTTGATTATCGAAAGAAGA
481        G  P  P  Q  C  D  Q  F  L  K  F  D  A  D  L  I  I  E  R  R

               1510      1520      1530      1540      1550      1560
1501      GAGGGTACTGATGTTTGTTACCCTGGAAAGTTCACAAACGAAGAGAGTCTTAGACAGATT
501        E  G  T  D  V  C  Y  P  G  K  F  T  N  E  E  S  L  R  Q  I

               1570      1580      1590      1600      1610      1620
1561      TTGAGAAGATCTGGTGGAATCGACAAAGAGTCCATGGGTTTTACTTACTCAGGTATTAGA
521        L  R  R  S  G  G  I  D  K  E  S  M  G  F  T  Y  S  G  I  R

               1630      1640      1650      1660      1670      1680
1621      ACCAACGGAACTACATCTGCATGCAGAAGATCCAATCCATCATTTTATGCTGAAATGAAG
541        T  N  G  T  T  S  A  C  R  R  S  N  P  S  F  Y  A  E  M  K

               1690      1700      1710      1720      1730      1740
1681      TGGTTGCTTTCCAACTCAGATAATGCTACCTTCCCTCAAATGACTAAGTCATATAGAAAC
561        W  L  L  S  N  S  D  N  A  T  F  P  Q  M  T  K  S  Y  R  N

               1750      1760      1770      1780      1790      1800
1741      CCAAGAAACAAGCCTGCCTTGATTACTTGGGGTGTTCATCACAGTGGATCTGCAACTGAG
581        P  R  N  K  P  A  L  I  T  W  G  V  H  H  S  G  S  A  T  E

               1810      1820      1830      1840      1850      1860
1801      CAAACAAAGCTTTACGGTTCTGGAGACAAGTTGATCACAGTTGGTAGTTCTAAGTATTTG
601        Q  T  K  L  Y  G  S  G  D  K  L  I  T  V  G  S  S  K  Y  L

               1870      1880      1890      1900      1910      1920
1861      CAGAGTTTTACCCCATCTCCTGGAGCCAGACCTCAAGTCAATGGACAGTCCGGTAGAATT
621        Q  S  F  T  P  S  P  G  A  R  P  Q  V  N  G  Q  S  G  R  I

               1930      1940      1950      1960      1970      1980
1921      GATTTCCATTGGTTGCTTTTGGATCCAAACGACACTGTTACTTTTACTTTTAACGGTGCT
641        D  F  H  W  L  L  L  D  P  N  D  T  V  T  F  T  F  N  G  A

               1990      2000      2010      2020      2030      2040
1981      TTTATTGCTCCTGACAGAGCATCTTTCTTTAGAGGAGAATCTATCGGAATCCAGTCATAA
661        F  I  A  P  D  R  A  S  F  F  R  G  E  S  I  G  I  Q  S  *


2041      TCTAGA
681        S  R
```

XhoI Site	1..6	6	==	Restriction Site
Kex2 Endopeptidase Signal	7..12	6	=>	misc_structure
AVIAN HA RBD V3.6	13..2052	2040	=>	CDS
HA RBD Subunit1 (H5) V3.4	13..678	666	=>	CDS
HA RBD Ser1/Gly4 Linker	679..711	33	=>	CDS
HA RBD Subunit2 (H5) V3.4	712..1365	654	=>	CDS
HA RBD Ser1/Gly4 Linker	1366..1398	33	=>	CDS
HA RBD Subunit 3 (H7) V3.4	1399..2037	639	=>	CDS
pSTOP	2038..2052	15	=>	CDS
XbaI site	2041..2046	3	==	Restriction Site

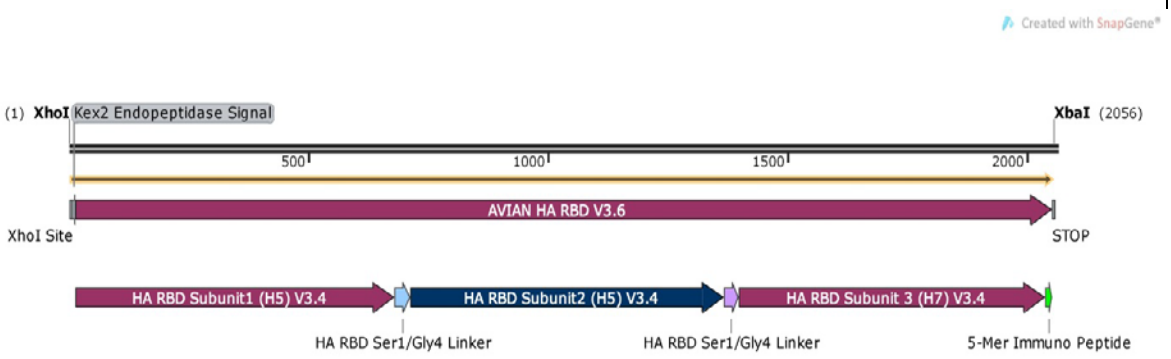
Table 13. DNA sequencing results for the Avian HA-RBD V3.6 version that was synthesized and codon optimized for *Pichia pastoris* expression (OptiGene

software by GeneScript Inc) in the vector pPicZalphaA. The presence of the XhoI, 5' restriction site facilitates cloning next to the Kex2 cleavage signal for N-terminal native expression. The XbaI, 3' restriction site; pSTOP, premature stop codon, RBD1, Receptor binding domain derived from H5; RBD2, Receptor binding domain derived from H5; RBD3, Receptor binding domain DNA sequence derived from H7; Pep5aa, an immune-stimulant penta peptide.

```
Translation of HA RBD V(3.6)(1-2061)
Universal code
Total amino acid number: 685, MW=76346
Max ORF starts at AA pos 1(may be DNA pos 1) for 684 AA(2052
bases),MW=76259
           10        20        30        40        50        60
1         CTCGAGAAGAGAAGTGTTAAGGGAGGTTTTATTTTGAAAGACTGTTCTGTCGCTGGTTGG
1          L  E  K  R  S  V  K  G  G  F  I  L  K  D  C  S  V  A  G  W

                  70        80        90       100       110       120
61        TTGCTTGGAAATCCTATGTGTGATGAGTTTTTGACTGCTCCAGAATGGTCCTACATTGTT
21         L  L  G  N  P  M  C  D  E  F  L  T  A  P  E  W  S  Y  I  V

                 130       140       150       160       170       180
121       GAGAAGGATAACCCATCAAATGACTTGTGTTACCCTGGTAACTTCGATAACTACGAAGAG
41         E  K  D  N  P  S  N  D  L  C  Y  P  G  N  F  D  N  Y  E  E

                 190       200       210       220       230       240
181       CTTAAGCATTTGATGTCTTCCACCAACCACATGGAAAGAATTCAAATCTTCCCAAGAAAT
61         L  K  H  L  M  S  S  T  N  H  M  E  R  I  Q  I  F  P  R  N

                 250       260       270       280       290       300
241       TCCTGGACTAACCATAATGCTTCAGCCGGTGTTTCAAGTGCTTGCCCTTTTAACGGTGGA
81         S  W  T  N  H  N  A  S  A  G  V  S  S  A  C  P  F  N  G  G

                 310       320       330       340       350       360
301       TCTTCCTTTTTCAGAAATGTTGTCTGGCTTATTAAGAAAAACGATGTTTACAGAACTTTG
101        S  S  F  F  R  N  V  V  W  L  I  K  K  N  D  V  Y  R  T  L

                 370       380       390       400       410       420
361       AAGAGAAACTACACCAACACTAATGTTGAGGATTTGCTTATTATCTGGGGTGTCCATCAC
121        K  R  N  Y  T  N  T  N  V  E  D  L  L  I  I  W  G  V  H  H

                 430       440       450       460       470       480
421       CCAAATGACGCTACTGAACAAACAAAGCTTTACCAGAACTTGAACACATACGTTAGTGTC
141        P  N  D  A  T  E  Q  T  K  L  Y  Q  N  L  N  T  Y  V  S  V

                 490       500       510       520       530       540
481       GGTACATCTACCTTGAACCAAAGATCAACTCCAACAATTGCCACTAGACCTCAAGTTAAT
161        G  T  S  T  L  N  Q  R  S  T  P  T  I  A  T  R  P  Q  V  N

                 550       560       570       580       590       600
541       GGTCAGAGAGGAAGAATGGAATTTTTCTGGACTATCTTGAAGTCCAACGATTCAATCTTT
181        G  Q  R  G  R  M  E  F  F  W  T  I  L  K  S  N  D  S  I  F

                 610       620       630       640       650       660
601       TTCGAGTCCACTGGTAATTTTATTGCCCCTGAATACGCATACAAGATCGTTAAGAAAGGA
201        F  E  S  T  G  N  F  I  A  P  E  Y  A  Y  K  I  V  K  K  G
```

```
                670       680       690       700       710       720
661       AAGTCCACTGTCATGAAATCTGGTGGAGGTGGAAGTGGTGGAGGTGGATCTCCACTTATT
221        K  S  T  V  M  K  S  G  G  G  G  S  G  G  G  G  S  P  L  I

                730       740       750       760       770       780
721       TTGAAGGATTGTTCTGTTGCTGGATGGTTGCTTGGTAATCCTATGTGCGACGAATTCTTG
241        L  K  D  C  S  V  A  G  W  L  L  G  N  P  M  C  D  E  F  L

                790       800       810       820       830       840
781       AACGTTCAAGAGTGGTCTTACATCGTCGAAAAGGATAACCCATCCAACGACTTGTGCTAT
261        N  V  Q  E  W  S  Y  I  V  E  K  D  N  P  S  N  D  L  C  Y

                850       860       870       880       890       900
841       CCAGGAAATTTTGACAATTATGAGGAGCTTAAACATTTGATGTCAAGTACTAACCACTTG
281        P  G  N  F  D  N  Y  E  E  L  K  H  L  M  S  S  T  N  H  L

                910       920       930       940       950       960
901       GAGAGAATTCAGATCTTCCCAAGAAACTCTTGGCCTAATCACGATTCTTCCTCAGGTGTT
301        E  R  I  Q  I  F  P  R  N  S  W  P  N  H  D  S  S  S  G  V

                970       980       990      1000      1010      1020
961       AGTTCTGCTTGTCCTTTTAACGGAGAATCCTCATTTTTCAGAAACGTTGTCTGGTTGATT
321        S  S  A  C  P  F  N  G  E  S  S  F  F  R  N  V  V  W  L  I

               1030      1040      1050      1060      1070      1080
1021      AAGAAAAATGACGTCTATAGAACCCTTAAGAGAAACTATACCAACACTAATGTCGAAGAT
341        K  K  N  D  V  Y  R  T  L  K  R  N  Y  T  N  T  N  V  E  D

               1090      1100      1110      1120      1130      1140
1081      TTGTTGATCATCTGGGGTGTTCATCACCCTAATGACGCAACCGTCCAAACTAAATTGTAC
361        L  L  I  I  W  G  V  H  H  P  N  D  A  T  V  Q  T  K  L  Y

               1150      1160      1170      1180      1190      1200
1141      CAGAACTTGAACACTTACGTTTCCGTCGGTACATCAACCTTGAACCAAAGATCTATTCCA
381        Q  N  L  N  T  Y  V  S  V  G  T  S  T  L  N  Q  R  S  I  P

               1210      1220      1230      1240      1250      1260
1201      ACAATCGCTACCAGACCACAGGTCAACGGTCAGAGAGGAAGAGTCGAGTTTTTCTGGACC
401        T  I  A  T  R  P  Q  V  N  G  Q  R  G  R  V  E  F  F  W  T

               1270      1280      1290      1300      1310      1320
1261      ATTTTGGAAAGTAACGATTCTATCTTTTTCGAGTCTACTGGTAACTTCATTGCTCCAGAA
421        I  L  E  S  N  D  S  I  F  F  E  S  T  G  N  F  I  A  P  E

               1330      1340      1350      1360      1370      1380
1321      TACGCCTATAAGGTTACAAAGAAAGGAAAGTCCGCTGTCATGAAGTCTGGAGGTGGAGGT
441        Y  A  Y  K  V  T  K  K  G  K  S  A  V  M  K  S  G  G  G  G

               1390      1400      1410      1420      1430      1440
1381      TCTGGAGGAGGTGGATCTCCTACCGATCTTGGTCAATGTGGATTGCTTGGTACTTTGATT
461        S  G  G  G  G  S  P  T  D  L  G  Q  C  G  L  L  G  T  L  I

               1450      1460      1470      1480      1490      1500
1441      GGTCCACCTCAATGCGACCAGTTTCTTAAATTCGATGCCGACTTGATTATCGAAAGAAGA
481        G  P  P  Q  C  D  Q  F  L  K  F  D  A  D  L  I  I  E  R  R
```

```
                1510        1520        1530        1540        1550        1560
1501        GAGGGTACTGATGTTTGTTACCCTGGAAAGTTCACAAACGAAGAGAGTCTTAGACAGATT
501          E   G   T   D   V   C   Y   P   G   K   F   T   N   E   E   S   L   R   Q   I

                1570        1580        1590        1600        1610        1620
1561        TTGAGAAGATCTGGTGGAATCGACAAAGAGTCCATGGGTTTTACTTACTCAGGTATTAGA
521          L   R   R   S   G   G   I   D   K   E   S   M   G   F   T   Y   S   G   I   R

                1630        1640        1650        1660        1670        1680
1621        ACCAACGGAACTACATCTGCATGCAGAAGATCCAATCCATCATTTTATGCTGAAATGAAG
541          T   N   G   T   T   S   A   C   R   R   S   N   P   S   F   Y   A   E   M   K

                1690        1700        1710        1720        1730        1740
1681        TGGTTGCTTTCCAACTCAGATAATGCTACCTTCCCTCAAATGACTAAGTCATATAGAAAC
561          W   L   L   S   N   S   D   N   A   T   F   P   Q   M   T   K   S   Y   R   N

                1750        1760        1770        1780        1790        1800
1741        CCAAGAAACAAGCCTGCCTTGATTACTTGGGGTGTTCATCACAGTGGATCTGCAACTGAG
581          P   R   N   K   P   A   L   I   T   W   G   V   H   H   S   G   S   A   T   E

                1810        1820        1830        1840        1850        1860
1801        CAAACAAAGCTTTACGGTTCTGGAGACAAGTTGATCACAGTTGGTAGTTCTAAGTATTTG
601          Q   T   K   L   Y   G   S   G   D   K   L   I   T   V   G   S   S   K   Y   L

                1870        1880        1890        1900        1910        1920
1861        CAGAGTTTTACCCCATCTCCTGGAGCCAGACCTCAAGTCAATGGACAGTCCGGTAGAATT
621          Q   S   F   T   P   S   P   G   A   R   P   Q   V   N   G   Q   S   G   R   I

                1930        1940        1950        1960        1970        1980
1921        GATTTCCATTGGTTGCTTTTGGATCCAAACGACACTGTTACTTTTACTTTTAACGGTGCT
641          D   F   H   W   L   L   L   D   P   N   D   T   V   T   F   T   F   N   G   A

                1990        2000        2010        2020        2030        2040
1981        TTTATTGCTCCTGACAGAGCATCTTTCTTTAGAGGAGAATCTATCGGAATCCAGTCAAAG
661          F   I   A   P   D   R   A   S   F   F   R   G   E   S   I   G   I   Q   S   K

                2050        2060
2041        TGGTGTGAGTGCTAATCTAGA
681          W   C   E   C   *   S   R
```



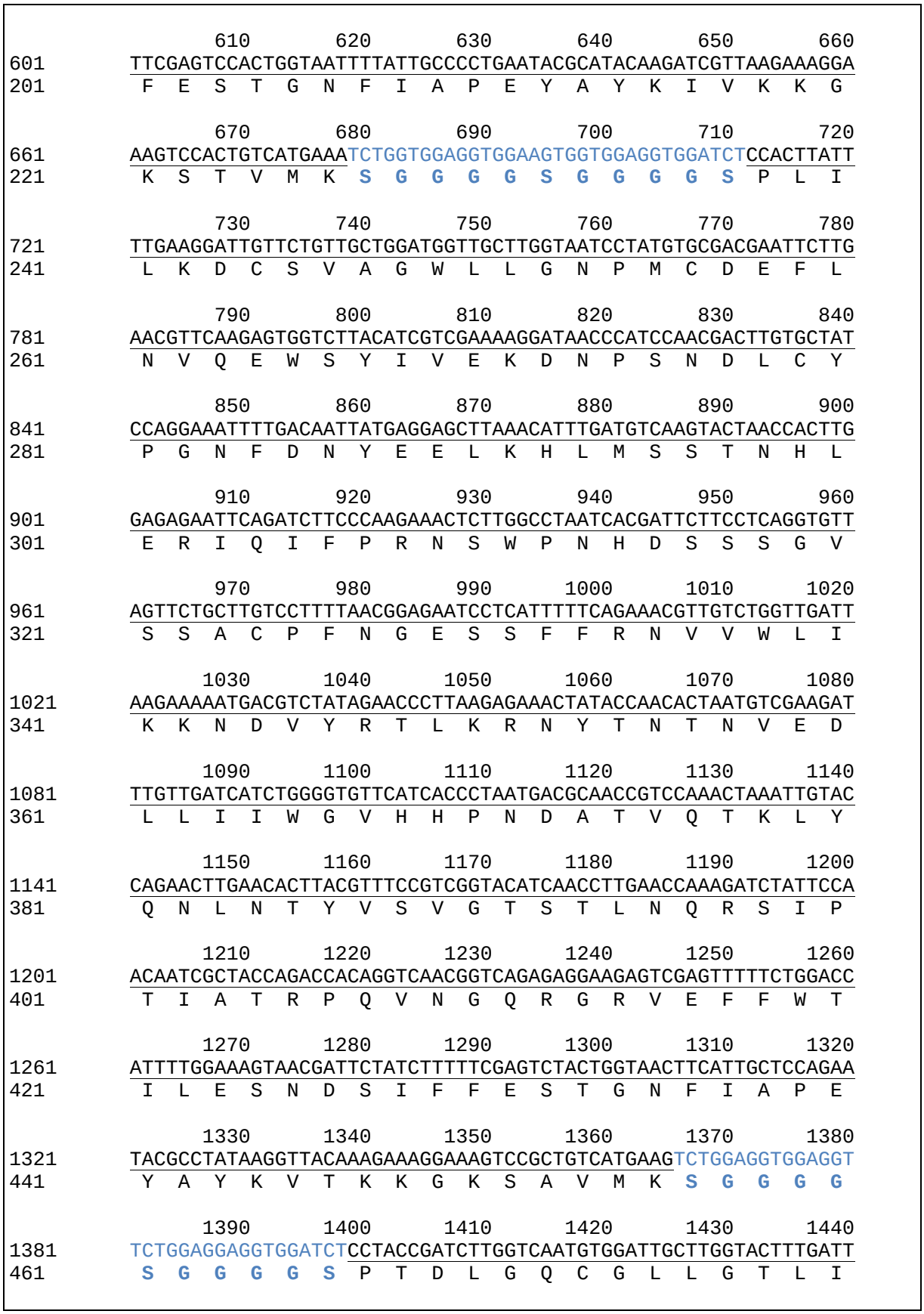
XhoI Site	1..6	6	==	misc_feature
Kex2 Endopeptidase Signal	7..12	6	=>	misc_structure

AVIAN HA RBD V3.6	13..2052	2040	=>	CDS
HA RBD Subunit1 (H5) V3.4	13..678	666	=>	CDS
HA RBD Ser1/Gly4 Linker	679..711	33	=>	CDS
HA RBD Subunit2 (H5) V3.4	712..1365	654	=>	CDS
HA RBD Ser1/Gly4 Linker	1366..1398	33	=>	CDS
HA RBD Subunit 3 (H7) V3.4	1399..2037	639	=>	CDS
5-Mer Immuno Peptide	2038..2052	15	=>	CDS
pSTOP	2053..2055	3	==	misc_feature

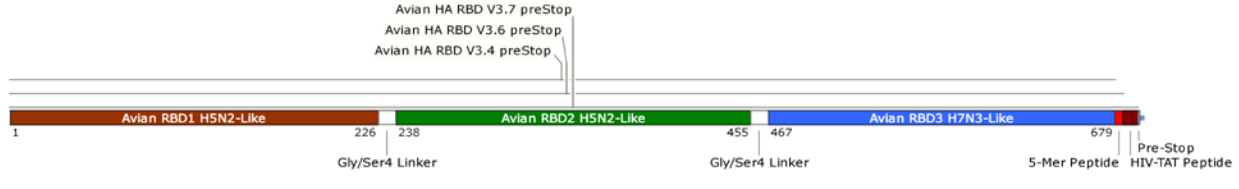
Table 14. DNA sequencing results for the Avian HA-RBD V3.7 version that was synthethized and codon optimized for *Pichia pastoris* expression (OptiGene software by GeneScript Inc) in the vector pPicZalphaA. (In Red) The XhoI, 5' and XbaI, 3' restriction sited were added to facilitate cloning. The pSTOP, premature stop codon, C-myc, detection epitope tag; 6His, Protein Recovery Tag; STOP, vector stop codon; RBD1, Receptor binding domain derived from H5; RBD2, Receptor binding domain derived from H5; RBD3, Receptor binding domain DNA sequence derived from H7; (In olive green) Pep5aa, immune-stimulant penta peptide; and Tat- pep , HIV-cell penetrating peptide are indicated.

Translation of **HA RBD V(3.7)**(1-2088)
Universal code
Total amino acid number: 694, MW=77666
Max ORF starts at AA pos 1(may be DNA pos 1) for 693 AA(2079 bases),MW=77579

```
          10        20        30        40        50        60
1         CTCGAGAAGAGAAGTGTTAAGGGAGGTTTTATTTTGAAAGACTGTTCTGTCGCTGGTTGG
1          L  E  K  R  S  V  K  G  G  F  I  L  K  D  C  S  V  A  G  W

          70        80        90       100       110       120
61        TTGCTTGGAAATCCTATGTGTGATGAGTTTTTGACTGCTCCAGAATGGTCCTACATTGTT
21         L  L  G  N  P  M  C  D  E  F  L  T  A  P  E  W  S  Y  I  V

         130       140       150       160       170       180
121       GAGAAGGATAACCCATCAAATGACTTGTGTTACCCTGGTAACTTCGATAACTACGAAGAG
41         E  K  D  N  P  S  N  D  L  C  Y  P  G  N  F  D  N  Y  E  E

         190       200       210       220       230       240
181       CTTAAGCATTTGATGTCTTCCACCAACCACATGGAAAGAATTCAAATCTTCCCAAGAAAT
61         L  K  H  L  M  S  S  T  N  H  M  E  R  I  Q  I  F  P  R  N

         250       260       270       280       290       300
241       TCCTGGACTAACCATAATGCTTCAGCCGGTGTTTCAAGTGCTTGCCCTTTTAACGGTGGA
81         S  W  T  N  H  N  A  S  A  G  V  S  S  A  C  P  F  N  G  G

         310       320       330       340       350       360
301       TCTTCCTTTTTCAGAAATGTTGTCTGGCTTATTAAGAAAAACGATGTTTACAGAACTTTG
101        S  S  F  F  R  N  V  V  W  L  I  K  K  N  D  V  Y  R  T  L

         370       380       390       400       410       420
361       AAGAGAAACTACACCAACACTAATGTTGAGGATTTGCTTATTATCTGGGGTGTCCATCAC
121        K  R  N  Y  T  N  T  N  V  E  D  L  L  I  I  W  G  V  H  H

         430       440       450       460       470       480
421       CCAAATGACGCTACTGAACAAACAAAGCTTTACCAGAACTTGAACACATACGTTAGTGTC
141        P  N  D  A  T  E  Q  T  K  L  Y  Q  N  L  N  T  Y  V  S  V

         490       500       510       520       530       540
481       GGTACATCTACCTTGAACCAAAGATCAACTCCAACAATTGCCACTAGACCTCAAGTTAAT
161        G  T  S  T  L  N  Q  R  S  T  P  T  I  A  T  R  P  Q  V  N

         550       560       570       580       590       600
541       GGTCAGAGAGGAAGAATGGAATTTTTCTGGACTATCTTGAAGTCCAACGATTCAATCTTT
181        G  Q  R  G  R  M  E  F  F  W  T  I  L  K  S  N  D  S  I  F
```


610 620 630 640 650 660
601 TTCGAGTCCACTGGTAATTTTATTGCCCCTGAATACGCATACAAGATCGTTAAGAAAGGA
201 F E S T G N F I A P E Y A Y K I V K K G
670 680 690 700 710 720
661 AAGTCCACTGTCATGAAATCTGGTGGAGGTGGAAGTGGTGGAGGTGGATCTCCACTTATT
221 K S T V M K S G G G G S G G G G S P L I
730 740 750 760 770 780
721 TTGAAGGATTGTTCTGTTGCTGGATGGTTGCTTGGTAATCCTATGTGCGACGAATTCTTG
241 L K D C S V A G W L L G N P M C D E F L
790 800 810 820 830 840
781 AACGTTCAAGAGTGGTCTTACATCGTCGAAAAGGATAACCCATCCAACGACTTGTGCTAT
261 N V Q E W S Y I V E K D N P S N D L C Y
850 860 870 880 890 900
841 CCAGGAAATTTTGACAATTATGAGGAGCTTAAACATTTGATGTCAAGTACTAACCACTTG
281 P G N F D N Y E E L K H L M S S T N H L
910 920 930 940 950 960
901 GAGAGAATTCAGATCTTCCCAAGAAACTCTTGGCCTAATCACGATTCTTCCTCAGGTGTT
301 E R I Q I F P R N S W P N H D S S S G V
970 980 990 1000 1010 1020
961 AGTTCTGCTTGTCCTTTTAACGGAGAATCCTCATTTTTCAGAAACGTTGTCTGGTTGATT
321 S S A C P F N G E S S F F R N V V W L I
1030 1040 1050 1060 1070 1080
1021 AAGAAAAATGACGTCTATAGAACCCTTAAGAGAAACTATACCAACACTAATGTCGAAGAT
341 K K N D V Y R T L K R N Y T N T N V E D
1090 1100 1110 1120 1130 1140
1081 TTGTTGATCATCTGGGGTGTTCATCACCCTAATGACGCAACCGTCCAAACTAAATTGTAC
361 L L I I W G V H H P N D A T V Q T K L Y
1150 1160 1170 1180 1190 1200
1141 CAGAACTTGAACACTTACGTTTCCGTCGGTACATCAACCTTGAACCAAAGATCTATTCCA
381 Q N L N T Y V S V G T S T L N Q R S I P
1210 1220 1230 1240 1250 1260
1201 ACAATCGCTACCAGACCACAGGTCAACGGTCAGAGAGGAAGAGTCGAGTTTTTCTGGACC
401 T I A T R P Q V N G Q R G R V E F F W T
1270 1280 1290 1300 1310 1320
1261 ATTTTGGAAAGTAACGATTCTATCTTTTTCGAGTCTACTGGTAACTTCATTGCTCCAGAA
421 I L E S N D S I F F E S T G N F I A P E
1330 1340 1350 1360 1370 1380
1321 TACGCCTATAAGGTTACAAAGAAAGGAAAGTCCGCTGTCATGAAGTCTGGAGGTGGAGGT
441 Y A Y K V T K K G K S A V M K S G G G G
1390 1400 1410 1420 1430 1440
1381 TCTGGAGGAGGTGGATCTCCTACCGATCTTGGTCAATGTGGATTGCTTGGTACTTTGATT
461 S G G G G S P T D L G Q C G L L G T L I

```
                1450       1460       1470       1480       1490       1500
1441        GGTCCACCTCAATGCGACCAGTTTCTTAAATTCGATGCCGACTTGATTATCGAAAGAAGA
481          G  P  P  Q  C  D  Q  F  L  K  F  D  A  D  L  I  I  E  R  R

                1510       1520       1530       1540       1550       1560
1501        GAGGGTACTGATGTTTGTTACCCTGGAAAGTTCACAAACGAAGAGAGTCTTAGACAGATT
501          E  G  T  D  V  C  Y  P  G  K  F  T  N  E  E  S  L  R  Q  I

                1570       1580       1590       1600       1610       1620
1561        TTGAGAAGATCTGGTGGAATCGACAAAGAGTCCATGGGTTTTACTTACTCAGGTATTAGA
521          L  R  R  S  G  G  I  D  K  E  S  M  G  F  T  Y  S  G  I  R

                1630       1640       1650       1660       1670       1680
1621        ACCAACGGAACTACATCTGCATGCAGAAGATCCAATCCATCATTTTATGCTGAAATGAAG
541          T  N  G  T  T  S  A  C  R  R  S  N  P  S  F  Y  A  E  M  K

                1690       1700       1710       1720       1730       1740
1681        TGGTTGCTTTCCAACTCAGATAATGCTACCTTCCCTCAAATGACTAAGTCATATAGAAAC
561          W  L  L  S  N  S  D  N  A  T  F  P  Q  M  T  K  S  Y  R  N

                1750       1760       1770       1780       1790       1800
1741        CCAAGAAACAAGCCTGCCTTGATTACTTGGGGTGTTCATCACAGTGGATCTGCAACTGAG
581          P  R  N  K  P  A  L  I  T  W  G  V  H  H  S  G  S  A  T  E

                1810       1820       1830       1840       1850       1860
1801        CAAACAAAGCTTTACGGTTCTGGAGACAAGTTGATCACAGTTGGTAGTTCTAAGTATTTG
601          Q  T  K  L  Y  G  S  G  D  K  L  I  T  V  G  S  S  K  Y  L

                1870       1880       1890       1900       1910       1920
1861        CAGAGTTTTACCCCATCTCCTGGAGCCAGACCTCAAGTCAATGGACAGTCCGGTAGAATT
621          Q  S  F  T  P  S  P  G  A  R  P  Q  V  N  G  Q  S  G  R  I

                1930       1940       1950       1960       1970       1980
1921        GATTTCCATTGGTTGCTTTTGGATCCAAACGACACTGTTACTTTTACTTTTAACGGTGCT
641          D  F  H  W  L  L  L  D  P  N  D  T  V  T  F  T  F  N  G  A

                1990       2000       2010       2020       2030       2040
1981        TTTATTGCTCCTGACAGAGCATCTTTCTTTAGAGGAGAATCTATCGGAATCCAGTCAAAG
661          F  I  A  P  D  R  A  S  F  F  R  G  E  S  I  G  I  Q  S  K

                2050       2060       2070       2080
2041        TGGTGTGAGTGCAGAAAGAAGAGAAGACAGAGAAGAAGATAATCTAGA
681          W  C  E  C  R  K  K  R  R  Q  R  R  R  *  S  R
```



XhoI Site 1..6 6 == misc_feature

Kex2 Endopeptidase Signal	7..12	6	=>	misc_structure
AVIAN HA RBD V3.6	13..2052	2040	=>	CDS
HA RBD Subunit1 (H5) V3.4	13..678	666	=>	CDS
HA RBD Ser1/Gly4 Linker	679..711	33	=>	CDS
HA RBD Subunit2 (H5) V3.4	712..1365	654	=>	CDS
HA RBD Ser1/Gly4 Linker	1366..1398	33	=>	CDS
HA RBD Subunit 3 (H7) V3.4	1399..2037	639	=>	CDS
5-Mer Immuno Peptide	2038..2052	15	=>	CDS
Tat-Peptide	2053..2079	27	=>	CDS
pSTOP	2080..2082	3	==	misc_feature

Table 15. Human HA-RBD V3.3 sequence that was codon optimized for *Pichia pastoris* expression using the software OptiGene by the company GeneScript Inc. (In Red) XhoI, 5' and XbaI, 3' restriction sited were added to facilitate cloning; a cleavage site KR for correct endopeptidase processing and native N-Terminal purification; pSTOP, premature stop codon (TAA, in light green) for C-terminal Native purification; RBD1, Receptor binding domain derived from H1 subtype; RBD2, Receptor binding domain derived from H3 subtype; RBD3, Receptor binding domain DNA sequence derived from Influenza B.

Translation of **HUMAN HA RBD V(3.3**(1-2145)
Universal code
Total amino acid number: 713, MW=78171
Max ORF starts at AA pos 1(may be DNA pos 1) for 712 AA(2136 bases),
MW=78084

```
                 10        20        30        40        50        60
1         CTCGAGAAAAGAGGTGTCGCTCCATTGCATTTGGGAAAGTGTAATATCGCCGGTTGGATC
1          L  E  K  R  G  V  A  P  L  H  L  G  K  C  N  I  A  G  W  I

                 70        80        90       100       110       120
61        CTTGGTAATCCTGAGTGTGAGTCATTGTCTACTGCTTCTTCCTGGAGTTACATTGTTGAA
21         L  G  N  P  E  C  E  S  L  S  T  A  S  S  W  S  Y  I  V  E

               130       140       150       160       170       180
121       ACACCTTCTGATAACGGTACCTGTTACCCAGGAGATTTCATCGACTATGAAGAGCTTAGA
41         T  P  S  D  N  G  T  C  Y  P  G  D  F  I  D  Y  E  E  L  R

               190       200       210       220       230       240
181       GAGCAATTGTCAAGTGTTTCTTCCTTTGAAAGATTCGAGATTTTTCCAAAGACATCAAGT
61         E  Q  L  S  S  V  S  S  F  E  R  F  E  I  F  P  K  T  S  S

               250       260       270       280       290       300
241       TGGCCTAATCATGACTCTAACAAAGGAGTTACCGCTGCCTGTCCTCACGCAGGTGCTAAG
81         W  P  N  H  D  S  N  K  G  V  T  A  A  C  P  H  A  G  A  K

               310       320       330       340       350       360
301       TCTTTCTACAAGAACTTGATCTGGCTTGTCAAGAAAGGAAATTCCTACCCAAAGTTGTCT
101        S  F  Y  K  N  L  I  W  L  V  K  K  G  N  S  Y  P  K  L  S

               370       380       390       400       410       420
361       AAGTCCTACATCAACGATAAGGGAAAAGAAGTTTTGGTCCTTTGGGGTATCCATCACCCA
121        K  S  Y  I  N  D  K  G  K  E  V  L  V  L  W  G  I  H  H  P

               430       440       450       460       470       480
421       TCAACTAGTGCTGATCAACAGTCATTGTACCAGAATGCCGACGCATATGTTTTCGTCGGT
141        S  T  S  A  D  Q  Q  S  L  Y  Q  N  A  D  A  Y  V  F  V  G

               490       500       510       520       530       540
481       TCTTCCAGATATTCTAAGAAGTTTAAGCCAGAAATCGCTATTAGACCTAAAGTTAGAGAT
161        S  S  R  Y  S  K  K  F  K  P  E  I  A  I  R  P  K  V  R  D

               550       560       570       580       590       600
541       CAAGAGGGTAGAATGAACTACTATTGGACTCTTGTCGAACCTGGAGACAAGATTACTTTC
181        Q  E  G  R  M  N  Y  Y  W  T  L  V  E  P  G  D  K  I  T  F
```

```
                610       620       630       640       650       660
601       GAGGCAACTGGTAACTTGGTTGTCCCAAGATACGCTTTTGCCATGGAAAGAAATGCTGGT
201        E  A  T  G  N  L  V  V  P  R  Y  A  F  A  M  E  R  N  A  G

                670       680       690       700       710       720
661       AGTGGAATTATCATTTCTGATGGTGGAGGTGGATCCGGTGGAGGTGGATCAGGAGAAATC
221        S  G  I  I  I  S  D  G  G  G  G  S  G  G  G  G  S  G  E  I

                730       740       750       760       770       780
721       TGCGATTCTCCTCATCAAATTCTTGACGGAGAGAATTGTACTTTGATTGATGCTTTGCTT
241        C  D  S  P  H  Q  I  L  D  G  E  N  C  T  L  I  D  A  L  L

                790       800       810       820       830       840
781       GGAGATCCACAATGCGACGGATTCCAGAACAAGAAATGGGACTTGTTTGTTGAGAGATCC
261        G  D  P  Q  C  D  G  F  Q  N  K  K  W  D  L  F  V  E  R  S

                850       860       870       880       890       900
841       AAGGCCTACTCAAATTGTTACCCATATGATGTTCCTGACTATGCATCATTGAGAAGTCTT
281        K  A  Y  S  N  C  Y  P  Y  D  V  P  D  Y  A  S  L  R  S  L

                910       920       930       940       950       960
901       GTCGCTTCAAGTGGTACATTGGAATTCAACAACGAGTCTTTTAATTGGACCGGTGTTACT
301        V  A  S  S  G  T  L  E  F  N  N  E  S  F  N  W  T  G  V  T

                970       980       990      1000      1010      1020
961       CAGAACGGAACCTCTTCCGCATGTATCAGAAGATCTAACAACTCCTTTTTCTCAAGATTG
321        Q  N  G  T  S  S  A  C  I  R  R  S  N  N  S  F  F  S  R  L

               1030      1040      1050      1060      1070      1080
1021      AATTGGCTTCACAGATTGAACTTCAAATACCCTGCCCTTAACGTTACAATGCCAAACAAC
341        N  W  L  H  R  L  N  F  K  Y  P  A  L  N  V  T  M  P  N  N

               1090      1100      1110      1120      1130      1140
1081      GAACAATTCGATAAGTTGTACATTTGGGGTGTCCATCACCCTGGAACTGACAAAGAGCAA
361        E  Q  F  D  K  L  Y  I  W  G  V  H  H  P  G  T  D  K  E  Q

               1150      1160      1170      1180      1190      1200
1141      ATTTTCTTGTACGCTCAGTCAAGTGGAAGAATCACAGTTAGTACCAAGAGATCTCAACAG
381        I  F  L  Y  A  Q  S  S  G  R  I  T  V  S  T  K  R  S  Q  Q

               1210      1220      1230      1240      1250      1260
1201      GCCGTCATCCCAAACATTGGTTCCAGACCTAGAATCAGAAACATCCCAAGTAGAATCTCT
401        A  V  I  P  N  I  G  S  R  P  R  I  R  N  I  P  S  R  I  S

               1270      1280      1290      1300      1310      1320
1261      ATCTACTGGACTATTGTTAAGCCTGGAGATATCTTGCTTATTAATTCCACTGGTAACTTG
421        I  Y  W  T  I  V  K  P  G  D  I  L  L  I  N  S  T  G  N  L

               1330      1340      1350      1360      1370      1380
1321      ATCGCCCCAAGAGGATACTTTAAGATCAGATCCGGTAAATCTTCCATTATGAGATCAGAT
441        I  A  P  R  G  Y  F  K  I  R  S  G  K  S  S  I  M  R  S  D

               1390      1400      1410      1420      1430      1440
1381      GCACCTATCGGTGGAGGTGGAAGTGGTGGAGGTGGATCTGGTACAAGAACCAGAGGAAAG
461        A  P  I  G  G  G  G  S  G  G  G  G  S  G  T  R  T  R  G  K
```

```
                1450       1460       1470       1480       1490       1500
1441       CTTTGTCCAGACTGCTTGAATTGTACTGATCTTGACGTTGCTTTGGGTAGACCTATGTGT
481         L   C   P   D   C   L   N   C   T   D   L   D   V   A   L   G   R   P   M   C

                1510       1520       1530       1540       1550       1560
1501       GTCGGAACTACACCATCCGCAAAGGCTTCAATTTTGCATGAAGTTAGACCTGTCACCTCT
501         V   G   T   T   P   S   A   K   A   S   I   L   H   E   V   R   P   V   T   S

                1570       1580       1590       1600       1610       1620
1561       GGATGCTTCCCAATTATGCACGATAGAACTAAAATCAGACAGCTTCCAAATTTGCTTAGA
521         G   C   F   P   I   M   H   D   R   T   K   I   R   Q   L   P   N   L   L   R

                1630       1640       1650       1660       1670       1680
1621       GGTTACGAAAACATCAGATTGTCTACTCAAAATGTTATCGACGCTGAGAAGGCCCCAGGT
541         G   Y   E   N   I   R   L   S   T   Q   N   V   I   D   A   E   K   A   P   G

                1690       1700       1710       1720       1730       1740
1681       GGACCTTACAGATTGGGTACCTCTGGATCCTGTCCTAACGCCACTTCTAAAATTGGTTTC
561         G   P   Y   R   L   G   T   S   G   S   C   P   N   A   T   S   K   I   G   F

                1750       1760       1770       1780       1790       1800
1741       TTTGCAACAATGGCCTGGGCAGTTCCAAAGGATAACTACAAAAATGCTACCAACCCTTTG
581         F   A   T   M   A   W   A   V   P   K   D   N   Y   K   N   A   T   N   P   L

                1810       1820       1830       1840       1850       1860
1801       ACTGTTGAAGTCCCATATATTTGCACTGAAGGAGAGGATCAAATCACAGTTTGGGGTTTT
601         T   V   E   V   P   Y   I   C   T   E   G   E   D   Q   I   T   V   W   G   F

                1870       1880       1890       1900       1910       1920
1861       CATTCTGACAATAAGACCCAAATGAAATCCTTGTATGGAGATTCAAACCCTCAGAAATTC
621         H   S   D   N   K   T   Q   M   K   S   L   Y   G   D   S   N   P   Q   K   F

                1930       1940       1950       1960       1970       1980
1921       ACTTCAAGTGCTAATGGAGTTACCACTCACTACGTCTCCCAAATTGGAGATTTTCCTGAC
641         T   S   S   A   N   G   V   T   T   H   Y   V   S   Q   I   G   D   F   P   D

                1990       2000       2010       2020       2030       2040
1981       CAGACTGAGGATGGTGGATTGCCACAATCAGGAAGAATTGTTGTCGACTACATGATGCAG
661         Q   T   E   D   G   G   L   P   Q   S   G   R   I   V   V   D   Y   M   M   Q

                2050       2060       2070       2080       2090       2100
2041       AAGCCAGGTAAAACTGGAACAATCGTTTATCAAAGAGGTGTCTTGTTGCCTCAAAAAGTC
681         K   P   G   K   T   G   T   I   V   Y   Q   R   G   V   L   L   P   Q   K   V

                2110       2120       2130       2140
2101       TGGTGTGCATCAGGTAGAAGTAAAGTCATCAAAGGATAATCTAGA
701         W   C   A   S   G   R   S   K   V   I   K   G   *
```

Table 16. RBD subunits and peptide composition for the HA-RBD genes versions used in the study. Versions V3.3, V3.4, V3.6, and V3.7 were synthetized and custom cloned in the pPicZalpha plasmid by GeneScript Corp.; whereas versions mutV3.3, mut V3.4, mutV3.6 and mutV3.7 were sequence modified by PCR and recloned in the plasmids pPicZalpha-A and pPicZ-B.

Avian Version	**Subunit 1**	**Subunit 2**	**Subunit 3**	**5-Mer peptide**	**TAT-peptide**	**Premature Stop Codon (Native C-Terminal)**	**c-Myc Tag**	**6-His Tag**	**Extra-cellular pPicZα-A**	**Intra-Cellular pPicZ-B**
V3.4	+	+	+	-	-	+	-	-	+	-
V3.6	+	+	+	+	-	+	-	-	+	-
V3.7	+	+	+	+	+	+	-	-	+	-
mutV3.4	+	+	+	-	-	-	+	+	+	+
mutV3.6	+	+	+	+	-	-	+	+	+	+
mutV3.7	+	+	+	+	+	-	+	+	+	+
Human Version	**Subunit 1**	**Subunit 2**	**Subunit 3**	**5-Mer peptide**	**TAT-peptide**	**Premature Stop Codon (Native C-Terminal)**	**c-Myc Tag**	**6-His Tag**	**Extra-cellular pPicZα-A**	**Intra-Cellular pPicZ-B**
V3.3	+	+	+	-	-	+	-	-	+	-
mutV3.3	+	+	+	-	-	-	+	+	+	+

3.3. Zeocin selection to obtain hyper-resistant *Pichia pastoris* clones carrying any of the HA-RBD gene versions.

A typical Pichia pastoris X-33 electro-transformation resulted in a highly efficient growth and recovery of hundreds of zeocine resistance colonies (**Figure 19 A**). Depending of the HA-RBD expression cassette being transformed, two to four Zeocine hyper-resistant clones at 1500 µL/mL antibiotic selection where recovered and stored, as observed and described (**Figure 19 B**). Thus, for a broader search of expressing clones, the panel of hyper-resistant cells at 1000 µL/mL where used for the expression analysis with the expectative to have better chances to induce higher expression level in individual clones than randomly picked ones (**Figure 19 C**).

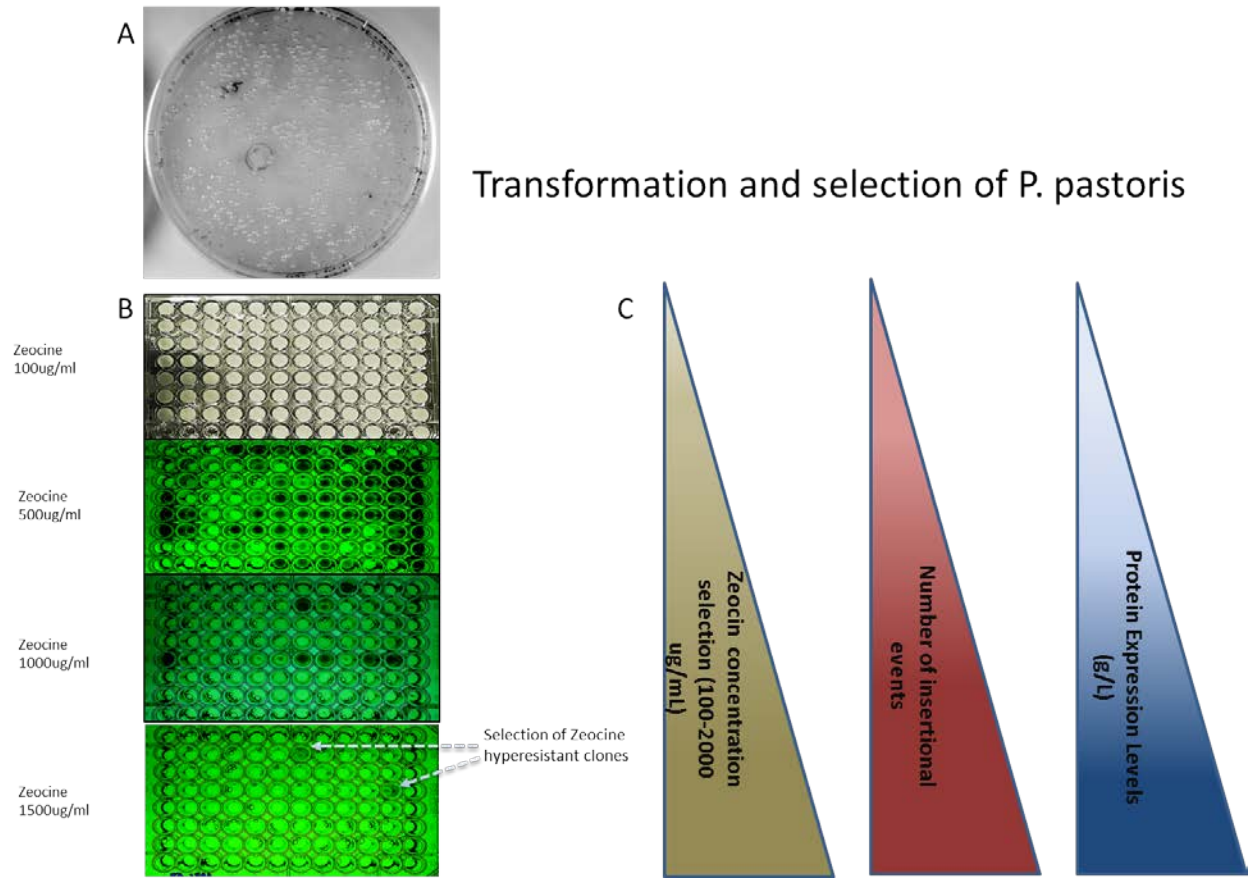


Figure 19. Antibiotic selection of the *Pichia pastoris* Zeocin resistant clones. A, *Pichia pastoris* X-33 cells were transformed with the HA RBD Version 3.4 cloned in the pPicZalpha-A plasmid and selected in YPDS-agar plates, containing 200 □g/□l of Zeocin. B, Isolated clones were selected using a tooth-pick and transferred individually to 96-well culture plates for further antibiotic hyperresistance selection. The cultures were replica plated in 96-well plates with different antibiotic concentrations, and those clones able to growth at the highest antibiotic concentration (dotted arrows) were selected for protein expression analysis. C, the expected results of applying this strategy is that cells hyper-resistant to zeocin will be carrying more gene copies in their genomes that will be translated into higher protein expression levels.

3.4. MASS-spectroscopy results from a protein band isolated from the avian HA RBD V3.4 – pPicZalpha-A Clone 1.1 identifies peptides derived from the Avian HA-RBD V3.4 protein sequence.

Pichia pastoris extracellular integrated **Avian HA-RBD V3.4 1.1** was observed to express a putative band under a methanol-induced 10 fold concentrated culture supernatant that migrates on a 12%SDS-PAGE gel between a molecular size of 100 KDa and 75 KDa (72 kDa expected). These bands (**Figure 20**) were recovered by DEAE anion exchange chromatography and further analyzed by Tandem Mass-Spectrometry to infer the peptide composition. The sequence generated by the MASCOT software compared directly to the Avian HA-RBD v3.4 protein sequence and Swiss-prot protein database and was used to generate peptide scores. Based on the total ion current score results in **Table 15**, it was possible to identify one protein corresponding to the Avian HA-RDB v3.4 sequence. From the total protein sequence length in (**Table 16**), 45% of protein sequence coverage to the Avian HA RBD v3.4 was observed, strongly indicating the expression of this protein in the isolated band. In a separate Mass-Spectroscopic analysis, a second purification by size exclusion chromatography in addition to the DEAE Anion exchange recovery (**Figure 21**) reconfirmed (**Table 17**) the presence of peptides corresponding to the Avian HA-RBD V3.4 sequence. However, this time there was only 29% protein sequence coverage when compared to the total protein length of the Avian HA-RBD 3.4 sequence (**Figure 22**).

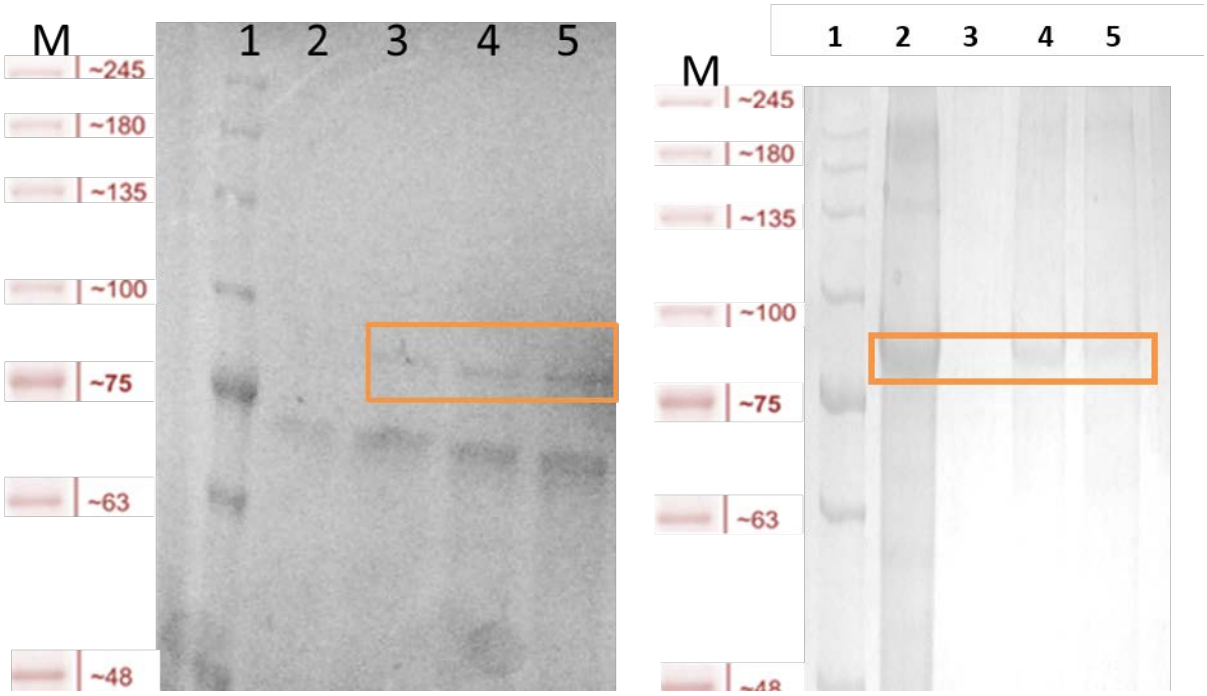


Figure 20. Expression analysis by SDS-PAGE of the Avian HA-RBD V3.4 Clone 1.1 from 48 h methanol-induced supernatants before (A) and after DEAE Anion Exchange chromatography purification (B). In A), Lane M and 1, Blueye protein standard; lane 2, mock induced culture with X-33 clones integrated with pPicZalphaA plasmid; lane 3-4, clone 1.1 integrated with Avian RDB V3.4 in pPicZalphaA, the orange box corresponds to a 75-100 KDa putative Avian HA-RBD V3.4 protein, intended for MASS-Spectroscopic analysis. In B, elution fractions in 300mM Tris-HCl pH 7.72 buffer, with the putative Avian HA-RBD V3.4; lane M and 1, Blueye protein standard, lane 2, second fraction collected and loaded in 300 mM Tris-HCl pH7.72 buffer; lane 3,4,5, correspond to the first, third and fourth, fractions collected at 300 mM Tris-HCl pH7.72 buffer. Second, third and fourth fractions were gel sliced and analysed by Mass-Spectroscopy.

Table 17. Calculated peptide sequences positively identified by Mass-Spectroscopy from the Avian HA-RBD V3.4 Clone 1.1. Highlighted in yellow, the calculated peptide sequences with a combination of a high score values and lowest expect value indicating more confidence in the match.

Query	Start	–	End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Peptide
5163	39	–	58	1202.5061	2402.9977	2402.9961	0.65	0	22	0.0083	K.DNPSNDLCYPGNFDNYEELK.H
5164	39	–	58	1202.5066	2402.9986	2402.9961	1.06	0	45	0.000062	K.DNPSNDLCYPGNFDNYEELK.H
3352	59	–	69	448.2039	1341.5899	1341.5918	-1.37	0	2	1.5	K.HLMSSTNHMER.I
3353	59	–	69	448.2039	1341.5899	1341.5918	-1.37	0	13	0.11	K.HLMSSTNHMER.I
609	102	–	108	436.2718	870.529	870.5327	-4.24	0	10	1.1	R.NVVWLIK.K
610	102	–	108	436.273	870.5315	870.5327	-1.37	0	7	2.4	R.NVVWLIK.K
611	102	–	108	436.2731	870.5317	870.5327	-1.22	0	10	0.83	R.NVVWLIK.K
612	102	–	108	436.2733	870.532	870.5327	-0.87	0	41	0.035	R.NVVWLIK.K
613	102	–	108	871.5402	870.5329	870.5327	0.26	0	27	0.21	R.NVVWLIK.K
614	102	–	108	436.2739	870.5332	870.5327	0.53	0	19	1.4	R.NVVWLIK.K
615	102	–	108	871.5406	870.5333	870.5327	0.69	0	23	0.24	R.NVVWLIK.K
616	102	–	108	436.274	870.5334	870.5327	0.81	0	43	0.019	R.NVVWLIK.K
1456	102	–	109	500.3205	998.6264	998.6277	-1.3	1	24	0.61	R.NVVWLIKK.N
1457	102	–	109	500.3207	998.6269	998.6277	-0.75	1	27	0.12	R.NVVWLIKK.N
1458	102	–	109	500.3208	998.6271	998.6277	-0.56	1	14	0.66	R.NVVWLIKK.N
1459	102	–	109	500.3213	998.6281	998.6277	0.47	1	23	0.85	R.NVVWLIKK.N
4991	146	–	164	1086.0581	2170.1017	2170.1018	-0.084	0	107	9.9E-09	K.LYQNLNTYVSVGTSTLNQR.S
4992	146	–	164	1086.0596	2170.1046	2170.1018	1.27	0	62	0.0000096	K.LYQNLNTYVSVGTSTLNQR.S

4993	146	–	164	1086.0599	2170.1053	2170.1018	1.61	0	102	4.3E-09	K.LYQNLNTYVSVGTSTLNQR.S
4994	146	–	164	1086.0604	2170.1063	2170.1018	2.05	0	81	0.0000002	K.LYQNLNTYVSVGTSTLNQR.S
4995	146	–	164	1086.5507	2171.0868	2170.1018	454	0	9	0.51	K.LYQNLNTYVSVGTSTLNQR.S
4996	146	–	164	1086.5509	2171.0873	2170.1018	454	0	31	0.0068	K.LYQNLNTYVSVGTSTLNQR.S
4997	146	–	164	1086.551	2171.0875	2170.1018	454	0	18	0.061	K.LYQNLNTYVSVGTSTLNQR.S
4998	146	–	164	1086.5516	2171.0887	2170.1018	455	0	40	0.0014	K.LYQNLNTYVSVGTSTLNQR.S
4194	165	–	179	542.628	1624.8621	1624.8645	-1.46	0	11	1.3	R.STPTIATRPQVNGQR.G
3614	180	–	190	714.3796	1426.7447	1426.7431	1.13	1	24	0.035	R.GRMEFFWTILK.S
3680	180	–	190	481.9203	1442.7392	1442.738	0.82	1	8	2.5	R.GRMEFFWTILK.S + Oxidation (M)
3681	180	–	190	722.3779	1442.7412	1442.738	2.18	1	10	0.52	R.GRMEFFWTILK.S + Oxidation (M)
5161	191	–	211	1200.5576	2399.1007	2399.0957	2.06	0	52	0.00018	K.SNDSIFFESTGNFIAPEYAYK.I
5162	191	–	211	1200.5582	2399.1019	2399.0957	2.57	0	39	0.00045	K.SNDSIFFESTGNFIAPEYAYK.I
3157	223	–	238	650.8461	1299.6777	1299.6783	-0.46	0	67	0.000017	K.SGGGGSGGGGSPLILK.D
3159	223	–	238	650.8465	1299.6784	1299.6783	0.11	0	106	0.000000011	K.SGGGGSGGGGSPLILK.D
3160	223	–	238	650.8466	1299.6786	1299.6783	0.2	0	91	0.000000094	K.SGGGGSGGGGSPLILK.D
3161	223	–	238	650.8467	1299.6789	1299.6783	0.48	0	54	0.00017	K.SGGGGSGGGGSPLILK.D
3162	223	–	238	650.8478	1299.681	1299.6783	2.08	0	18	0.21	K.SGGGGSGGGGSPLILK.D
3275	288	–	298	442.2183	1323.633	1323.6353	-1.74	0	5	1.1	K.HLMSSTNHLER.I
3276	288	–	298	442.2184	1323.6334	1323.6353	-1.46	0	16	0.1	K.HLMSSTNHLER.I
3277	288	–	298	442.2184	1323.6335	1323.6353	-1.39	0	21	0.36	K.HLMSSTNHLER.I
5395	305	–	330	954.4016	2860.1828	2859.1943	346	0	8	0.18	R.NSWPNHDSSSGVSSACPFNGESSFFR.N
5396	305	–	330	954.4019	2860.1839	2859.1943	346	0	25	0.01	R.NSWPNHDSSSGVSSACPFNGESSFFR.N

4236	394	–	408	819.4578	1636.9011	1636.9009	0.12	0	2	1.1	R.SIPTIATRPQVNGQR.G
4238	394	–	408	546.6412	1636.9017	1636.9009	0.49	0	19	0.55	R.SIPTIATRPQVNGQR.G
4252	394	–	408	546.9699	1637.8879	1636.9009	603	0	9	1.2	R.SIPTIATRPQVNGQR.G
2038	487	–	495	546.2895	1090.5645	1090.5659	-1.2	0	36	0.012	K.FDADLIIER.R
2039	487	–	495	546.2903	1090.5661	1090.5659	0.25	0	65	0.00014	K.FDADLIIER.R
2040	487	–	495	546.2906	1090.5666	1090.5659	0.7	0	53	0.0023	K.FDADLIIER.R
2885	487	–	496	416.563	1246.667	1246.667	0.065	1	16	0.85	K.FDADLIIERR.E
2888	487	–	496	624.3413	1246.6681	1246.667	0.88	1	11	0.68	K.FDADLIIERR.E
2889	487	–	496	624.342	1246.6695	1246.667	2.06	1	5	1.3	K.FDADLIIERR.E
5087	496	–	514	565.2687	2257.0456	2257.0433	1.01	2	15	0.15	R.REGTDVCYPGKFTNEESLR.Q
5088	496	–	514	753.356	2257.046	2257.0433	1.2	2	16	0.057	R.REGTDVCYPGKFTNEESLR.Q
2229	497	–	506	563.2469	1124.4792	1124.4808	-1.43	0	10	0.31	R.EGTDVCYPGK.F
2230	497	–	506	563.2476	1124.4807	1124.4808	-0.13	0	0	1.4	R.EGTDVCYPGK.F
4912	497	–	514	701.3212	2100.9417	2100.9422	-0.27	1	9	0.14	R.EGTDVCYPGKFTNEESLR.Q
4913	497	–	514	701.3213	2100.942	2100.9422	-0.094	1	22	0.015	R.EGTDVCYPGKFTNEESLR.Q
4914	497	–	514	701.3218	2100.9437	2100.9422	0.69	1	37	0.00038	R.EGTDVCYPGKFTNEESLR.Q
4915	497	–	514	1051.4796	2100.9447	2100.9422	1.16	1	25	0.041	R.EGTDVCYPGKFTNEESLR.Q
4921	497	–	514	1051.9803	2101.9461	2100.9422	478	1	11	0.29	R.EGTDVCYPGKFTNEESLR.Q
1436	507	–	514	498.2425	994.4704	994.472	-1.59	0	8	2.1	K.FTNEESLR.Q
1437	507	–	514	498.2429	994.4712	994.472	-0.79	0	30	0.042	K.FTNEESLR.Q
1438	507	–	514	498.2434	994.4722	994.472	0.25	0	20	0.82	K.FTNEESLR.Q
2758	515	–	525	613.3594	1224.7042	1224.6939	8.44	2	1	4.1	R.QILRRSGGIDK.E + Gln->pyro-Glu (N-term Q)

4819	519	–	536	654.3232	1959.9477	1959.9472	0.24	2	48	0.00024	R.RSGGIDKESMGFTYSGIR.T
4833	519	–	536	659.6543	1975.9411	1975.9422	-0.56	2	8	1.5	R.RSGGIDKESMGFTYSGIR.T + Oxidation (M)
4835	519	–	536	659.6545	1975.9418	1975.9422	-0.19	2	29	0.012	R.RSGGIDKESMGFTYSGIR.T + Oxidation (M)
4839	519	–	536	659.6553	1975.944	1975.9422	0.92	2	59	0.000014	R.RSGGIDKESMGFTYSGIR.T + Oxidation (M)
4612	520	–	536	902.9308	1803.847	1803.8461	0.49	1	24	0.0066	R.SGGIDKESMGFTYSGIR.T
4613	520	–	536	602.2899	1803.8479	1803.8461	0.99	1	40	0.00053	R.SGGIDKESMGFTYSGIR.T
4629	520	–	536	607.6207	1819.8404	1819.8411	-0.39	1	42	0.0016	R.SGGIDKESMGFTYSGIR.T + Oxidation (M)
4630	520	–	536	910.9279	1819.8412	1819.8411	0.055	1	16	0.15	R.SGGIDKESMGFTYSGIR.T + Oxidation (M)
4631	520	–	536	607.6212	1819.8418	1819.8411	0.41	1	5	3.4	R.SGGIDKESMGFTYSGIR.T + Oxidation (M)
4632	520	–	536	607.6213	1819.842	1819.8411	0.52	1	13	0.12	R.SGGIDKESMGFTYSGIR.T + Oxidation (M)
4633	520	–	536	910.9283	1819.842	1819.8411	0.52	1	24	0.051	R.SGGIDKESMGFTYSGIR.T + Oxidation (M)
4634	520	–	536	910.9288	1819.843	1819.8411	1.06	1	49	0.00005	R.SGGIDKESMGFTYSGIR.T + Oxidation (M)
4635	520	–	536	607.6219	1819.8438	1819.8411	1.52	1	49	0.000093	R.SGGIDKESMGFTYSGIR.T + Oxidation (M)
2882	526	–	536	624.2897	1246.5648	1246.5652	-0.34	0	29	0.049	K.ESMGFTYSGIR.T
2977	526	–	536	632.2868	1262.5591	1262.5601	-0.86	0	22	0.016	K.ESMGFTYSGIR.T + Oxidation (M)
2978	526	–	536	632.2871	1262.5597	1262.5601	-0.37	0	23	0.0065	K.ESMGFTYSGIR.T + Oxidation (M)
3299	546	–	556	665.3163	1328.618	1328.6183	-0.2	1	42	0.0018	R.RSNPSFYAEMK.W
3300	546	–	556	665.3163	1328.6181	1328.6183	-0.11	1	16	0.17	R.RSNPSFYAEMK.W
3301	546	–	556	665.3167	1328.6189	1328.6183	0.44	1	48	0.0014	R.RSNPSFYAEMK.W
3374	546	–	556	449.2111	1344.6115	1344.6132	-1.24	1	11	0.34	R.RSNPSFYAEMK.W + Oxidation (M)
3375	546	–	556	673.313	1344.6115	1344.6132	-1.24	1	3	1	R.RSNPSFYAEMK.W + Oxidation (M)
3376	546	–	556	673.3137	1344.6129	1344.6132	-0.24	1	20	0.4	R.RSNPSFYAEMK.W + Oxidation (M)

2509	547	–	556	587.2653	1172.5161	1172.5172	-0.94	0	16	0.11	R.SNPSFYAEMK.W
2510	547	–	556	587.2656	1172.5167	1172.5172	-0.42	0	50	0.00016	R.SNPSFYAEMK.W
2511	547	–	556	1173.5244	1172.5171	1172.5172	-0.04	0	15	0.27	R.SNPSFYAEMK.W
2512	547	–	556	587.2661	1172.5176	1172.5172	0.32	0	55	0.00012	R.SNPSFYAEMK.W
2513	547	–	556	1173.525	1172.5177	1172.5172	0.48	0	15	0.4	R.SNPSFYAEMK.W
2514	547	–	556	587.2667	1172.5188	1172.5172	1.36	0	17	0.054	R.SNPSFYAEMK.W
2580	547	–	556	595.2623	1188.5101	1188.5121	-1.68	0	31	0.0063	R.SNPSFYAEMK.W + Oxidation (M)
2581	547	–	556	595.2629	1188.5113	1188.5121	-0.65	0	51	0.00032	R.SNPSFYAEMK.W + Oxidation (M)
2583	547	–	556	595.263	1188.5114	1188.5121	-0.55	0	4	0.94	R.SNPSFYAEMK.W + Oxidation (M)
2584	547	–	556	595.2632	1188.5119	1188.5121	-0.14	0	11	1.2	R.SNPSFYAEMK.W + Oxidation (M)
5091	579	–	599	566.296	2261.1547	2261.1553	-0.24	0	17	0.48	R.NKPALITWGVHHSGSATEQTK.L
5092	579	–	599	566.296	2261.155	2261.1553	-0.13	0	27	0.017	R.NKPALITWGVHHSGSATEQTK.L
3931	600	–	614	508.9466	1523.8179	1523.8195	-1.09	1	12	0.23	K.LYGSGDKLITVGSSK.Y
3932	600	–	614	762.9166	1523.8186	1523.8195	-0.62	1	35	0.011	K.LYGSGDKLITVGSSK.Y
3934	600	–	614	508.947	1523.8191	1523.8195	-0.3	1	14	0.34	K.LYGSGDKLITVGSSK.Y
3935	600	–	614	762.917	1523.8194	1523.8195	-0.056	1	9	0.22	K.LYGSGDKLITVGSSK.Y
3936	600	–	614	508.9471	1523.8196	1523.8195	0.056	1	32	0.0078	K.LYGSGDKLITVGSSK.Y
3937	600	–	614	508.9472	1523.8197	1523.8195	0.11	1	54	0.00056	K.LYGSGDKLITVGSSK.Y
3939	600	–	614	762.9177	1523.8209	1523.8195	0.9	1	82	0.00000085	K.LYGSGDKLITVGSSK.Y
5078	615	–	635	749.7138	2246.1196	2246.1192	0.16	0	5	1.2	K.YLQSFTPSPGARPQVNGQSGR.I
5081	615	–	635	750.0419	2247.1038	2246.1192	438	0	10	0.77	K.YLQSFTPSPGARPQVNGQSGR.I
5083	615	–	635	750.0428	2247.1067	2246.1192	440	0	7	1.1	K.YLQSFTPSPGARPQVNGQSGR.I

3530	663	–	675	699.854	1397.6935	1397.6939	-0.34	1	28	0.019	R.ASFFRGESIGIQS.-
3531	663	–	675	699.8549	1397.6952	1397.6939	0.89	1	31	0.0075	R.ASFFRGESIGIQS.-
3532	663	–	675	699.8553	1397.6961	1397.6939	1.59	1	62	0.000019	R.ASFFRGESIGIQS.-

Table 18. Details of Avian HA-RBD V3.4 Peptides Matched to Observed Spectra. In red are shown the calculated peptides from (Table 14) which comprised a partial coverage of 45% from the total Avian HA-RBD V3.4 protein sequence.

1	SVKGGFILKD	CSVAGWLLGN	PMCDEFLTAP	EWSYIVEKDN	PSNDLCYPGN
51	FDNYEELKHL	MSSTNHMERI	QIFPRNSWTN	HNASAGVSSA	CPFNGGSSFF
101	RNVVWLIKKN	DVYRTLKRNY	TNTNVEDLLI	IWGVHHPNDA	TEQTKLYQNL
151	NTYVSVGTST	LNQRSTPTIA	TRPQVNGQRG	RMEFFWTILK	SNDSIFFEST
201	GNFIAPEYAY	KIVKKGKSTV	MKSGGGGSGG	GGSPLILKDC	SVAGWLLGNP
251	MCDEFLNVQE	WSYIVEKDNP	SNDLCYPGNF	DNYEELKHLM	SSTNHLERIQ
301	IFPRNSWPNH	DSSSGVSSAC	PFNGESSFFR	NVVWLIKKND	VYRTLKRNYT
351	NTNVEDLLII	WGVHHPNDAT	VQTKLYQNLN	TYVSVGTSTL	NQRSIPTIAT
401	RPQVNGQRGR	VEFFWTILES	NDSIFFESTG	NFIAPEYAYK	VTKKGKSAVM
451	KSGGGGSGGG	GSPTDLGQCG	LLGTLIGPPQ	CDQFLKFDAD	LIIERREGTD
501	VCYPGKFTNE	ESLRQILRRS	GGIDKESMGF	TYSGIRTNGT	TSACRRSNPS
551	FYAEMKWLLS	NSDNATFPQM	TKSYRNPRNK	PALITWGVHH	SGSATEQTKL
601	YGSGDKLITV	GSSKYLQSFT	PSPGARPQVN	GQSGRIDFHW	LLLDPNDTVT
651	FTFNGAFIAP	DRASFFRGES	IGIQS		

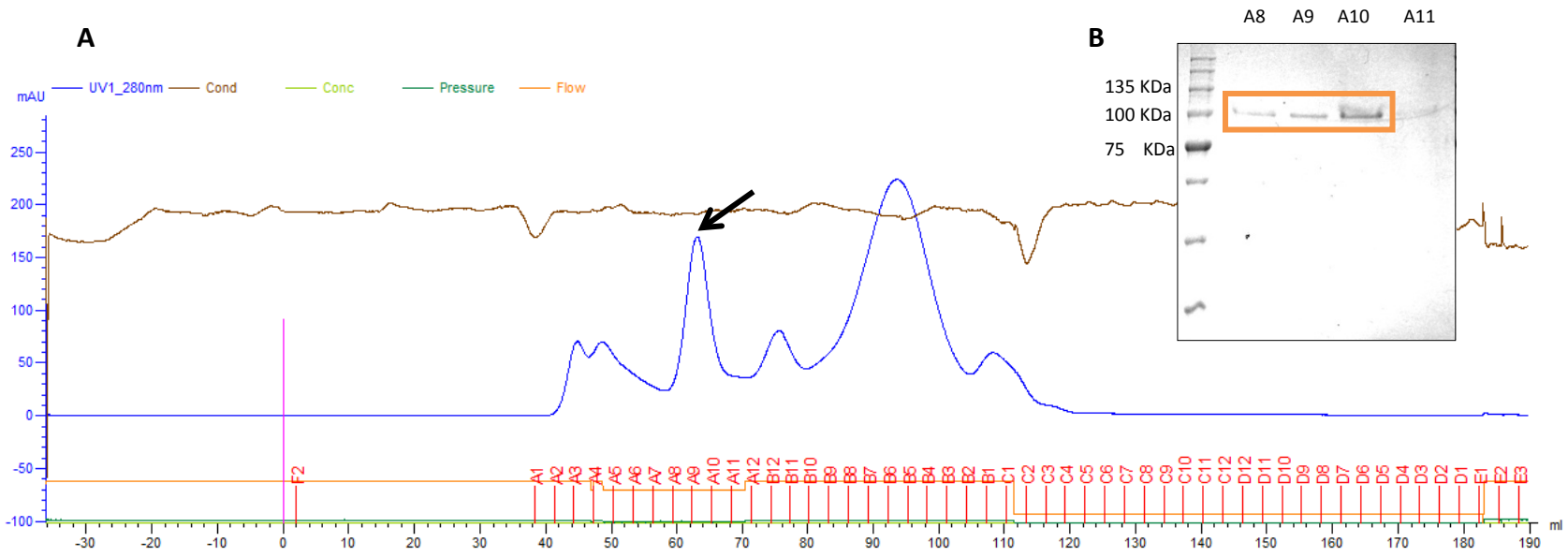


Figure 21. Protein purification of Avian HA-RBD V3.4 by size exclusion chromatography. The characterization of the eluted fractions following separation on DEAE Anion exchange chromatography. The sample fractions were pooled and concentrated to 1.6 ml using a 30 MWCO ultrafiltration device (Centricon) and then loaded by injection onto an AKTA system fitted with a Superdex 75 pq Hi Load 16-200 (10,000-600,000 MW) size exclusion column. Fractions were eluted using PBS, pH 7.4 as the isocratic mobile phase and buffer exchange solution. Arrow in A) indicates the protein peak fractions selected (A8,A9,A10) for Mass-Spec Analysis and N-terminal sequencing as observed in B), after SDS-PAGE analysis (orange box).

Table 19. Calculated peptide sequences positively identified by Mass-Spectroscopic analysis from the Avian HA RBD V3.4 Clone 1.1. As in Table 15, the calculated peptide sequences were determined using a combination of the high score values and the lowest expect value to indicate more confidence in the match analysis. The sample was purified by DEAE Anion Exchange chromatography followed by Size exclusion chromatography and SDS-PAGE and the protein gel slice isolated and characterized by mass spectroscopy.

Query	Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide
2394	39 – 58	1202.5076	2403.0006	2402.9961	1.87	0	45	6.5e-05	1	U	K.DNPSNDLCYPGNFDNYEELK.H
253	102 – 108	436.2737	870.5328	870.5327	0.11	0	13	1.2	2	U	R.NVVWLIK.K
254	102 – 108	871.5403	870.5330	870.5327	0.33	0	18	0.51	1	U	R.NVVWLIK.K
2206	146 – 164	1086.0598	2170.1051	2170.1018	1.49	0	100	3.6e-09	1	U	K.LYQNLNTYVSVGTSTLNQR.S
991	182 – 190	607.8187	1213.6229	1213.6205	1.94	0	18	0.33	1	U	R.MEFFWTILK.S
1018	182 – 190	615.8153	1229.6161	1229.6155	0.49	0	16	0.73	1	U	R.MEFFWTILK.S + Oxidation (M)
1124	223 – 238	650.8472	1299.6798	1299.6783	1.14	0	21	0.047	1	U	K.SGGGGSGGGGSPLILK.D
3075	305 – 330	954.4016	2860.1830	2859.1943	346	0	11	0.087	1	U	R.NSWPNHDSSSGVSSACPFNGESSFFR.N
221	444 – 451	864.4937	863.4864	863.4899	-3.96	2	2	3	8	U	K.KGKSAVMK.S + Oxidation (M)
778	487 – 495	546.2902	1090.5659	1090.5659	0.034	0	46	0.012	1	U	K.FDADLIIER.R
2001	497 – 514	701.3223	2100.9450	2100.9422	1.30	1	28	0.0044	1	U	R.EGTDVCYPGKFTNEESLR.Q
1864	519 – 536	659.6545	1975.9416	1975.9422	-0.28	2	10	0.44	1	U	R.RSGGIDKESMGFTYSGIR.T + Oxidation (M)
1670	520 – 536	902.9302	1803.8458	1803.8461	-0.20	1	16	0.13	1	U	R.SGGIDKESMGFTYSGIR.T
1671	520 – 536	602.2897	1803.8472	1803.8461	0.58	1	25	0.05	1	U	R.SGGIDKESMGFTYSGIR.T
1694	520 – 536	607.6219	1819.8438	1819.8411	1.52	1	29	0.0041	1	U	R.SGGIDKESMGFTYSGIR.T + Oxidation (M)
920	547 – 556	587.2662	1172.5178	1172.5172	0.52	0	15	0.12	1	U	R.SNPSFYAEMK.W
2313	579 – 599	566.2964	2261.1564	2261.1553	0.52	0	18	0.093	1	U	R.NKPALITWGVHHSGSATEQTK.L
1361	600 – 614	762.9166	1523.8186	1523.8195	-0.62	1	14	0.19	1	U	K.LYGSGDKLITVGSSK.Y
1362	600 – 614	508.9468	1523.8187	1523.8195	-0.55	1	13	0.27	1	U	K.LYGSGDKLITVGSSK.Y

A

1	SVKGGFILKD	CSVAGWLLGN	PMCDEFLTAP	EWSYIVEK**DN**	**PSNDLCYPGN**
51	**FDNYEELK**HL	MSSTNHMERI	QIFPRNSWTN	HNASAGVSSA	CPFNGGSSFF
101	R**NVVWLIK**KN	DVYRTLKRNY	TNTNVEDLLI	IWGVHHPNDA	TEQTK**LYQNL**
151	**NTYVSVGTST**	**LNQR**STPTIA	TRPQVNGQRG	R**MEFFWTILK**	SNDSIFFEST
201	GNFIAPEYAY	KIVKKGKSTV	MK**SGGGGSGG**	**GGSPLILK**DC	SVAGWLLGNP
251	MCDEFLNVQE	WSYIVEKDNP	SNDLCYPGNF	DNYEELKHLM	SSTNHLERIQ
301	IFPR**NSWPNH**	**DSSSGVSSAC**	**PFNGESSFFR**	NVVWLIKKND	VYRTLKRNYT
351	NTNVEDLLII	WGVHHPNDAT	VQTKLYQNLN	TYVSVGTSTL	NQRSIPTIAT
401	RPQVNGQRGR	VEFFWTILES	NDSIFFESTG	NFIAPEYAYK	VTK**KGKSAVM**
451	**K**SGGGGSGGG	GSPTDLGQCG	LLGTLIGPPQ	CDQFLK**FDAD**	**LIIER**R**EGTD**
501	**VCYPGKFTNE**	**ESLR**QILR**RS**	**GGIDKESMGF**	**TYSGIR**TNGT	TSACRR**SNPS**
551	**FYAEMK**WLLS	NSDNATFPQM	TKSYRNPR**NK**	**PALITWGVHH**	**SGSATEQTKL**
601	**YGSGDKLITV**	**GSSK**YLQSFT	PSPGARPQVN	GQSGRIDFHW	LLLDPNDTVT
651	FTFNGAFIAP	DRASFFRGES	IGIQS		

B

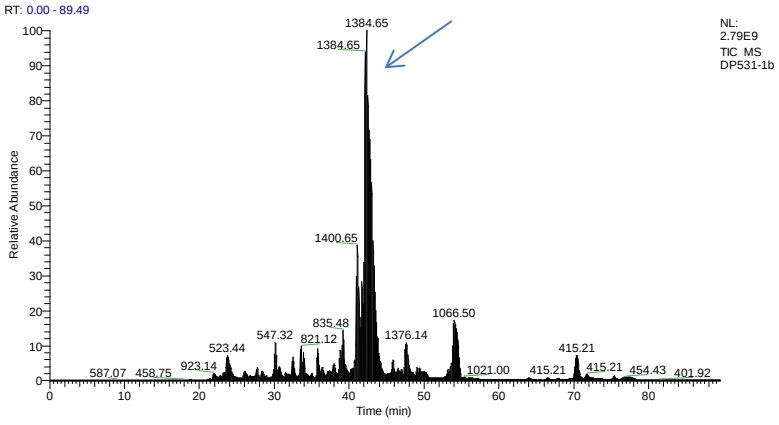


Figure 22. Details of Avian HA-RBD V3.4 Peptides Matched to Observed Spectra. In A), Highlighted in bold, are shown the calculated peptides from the MASCOT spectra analysis observed in B), and from Table 16, which comprised 29% of partial coverage from the total Avian HA RBD V3.4 protein sequence. The arrow indicates the spectra result in a major peak for which it was not possible to identify the type of chemistry by the method applied for that particular signal.

3.5. Immunoblot and N-terminal sequencing analysis fail to identify Avian HA RBD HA3.4 from positive MASS-Spectroscopy samples isolated by Anion exchange and Size exclusion chromatography.

Immunodetection analysis using ferret polyclonals raised against AIV H7N2, AIV H5N1 failed to detect the purified protein at the expected molecular size (~72KDa-100KDa). A positive signal was observed for the sample purified by Anion Chromatography between 180 KDa and 245 KDa of the molecular standard, suggesting a reactive band at a higher molecular weight under denaturing and reducing conditions (**Figure 24**, A Lane 6 and B Lane 6). On the other hand, the N-terminal analysis of the putative Avian RBD HA V3.4 protein identified derivatized amino acids with the sequence N-(E/A)-Y-V-P-Q-(E-I)-P (**Table 18**) that was in contrast with the expected sequence S-V-K-G-G-F-I-L. Single and clear peaks were not observed in the chromatogram results (Figure 21), as was expected from the sequence of the purified protein.

Table 20. N-Terminal sequencing analysis for the putative band identified by Mass-Spectroscopy as the Avian HA-RBD V3.3 derived from Clone 1.1. The analysis resulted in an derivatized N-Terminal Sequence: (E/A)-Y-V-P-Q-(E-I)-P which was not found in the expected protein sequence residues (SVKGGFIL).

Customer Code	Sample Name	Sample Code	Analysis Date
MitomaFD	HA 34	021015CA	2015-02-10

Residue:	Amino Acid:	Comments:
Residue 1		Possible amino acids for Residue 1: E-Glu; A-Ala
Residue 2	Y - Tyr	
Residue 3	V - Val	
Residue 4	P - Pro	
Residue 5	Q - Gln	
Residue 6		Possible amino acids for Residue 6: E-Glu; I-Ile
Residue 7	P - Pro	

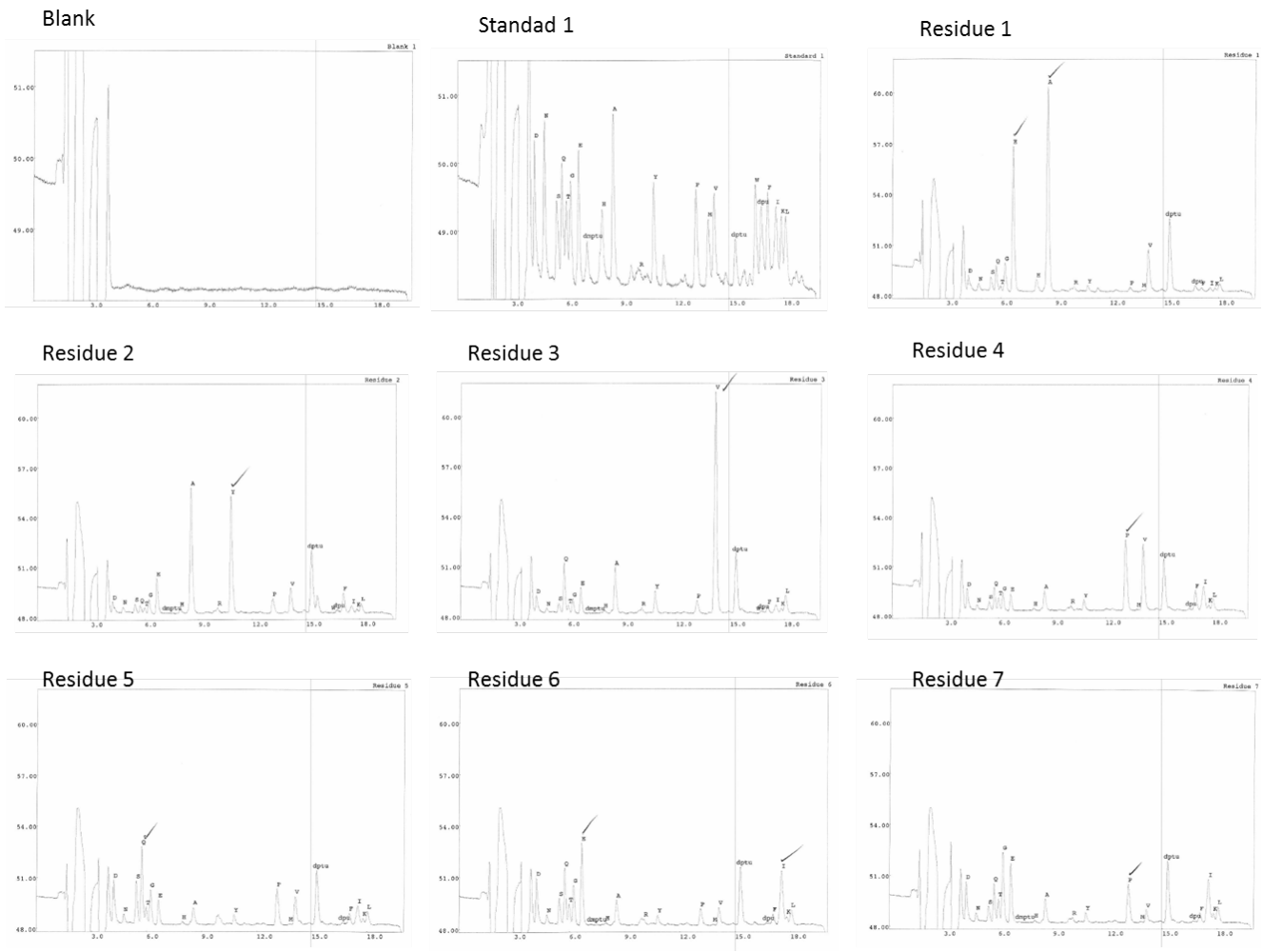


Figure 23. N- Terminal Analysis of the putative Avian HA-RBD V3.4. Chromatograms of the Derivatized amino group obtained from Edman-degradation of the N-terminal peptide. These chromatograms did not result in clear derivatized peaks, indicating a mixed population of proteins (which complement the Mass Spectroscopy-Results) or multiple proteolyzed peptides.

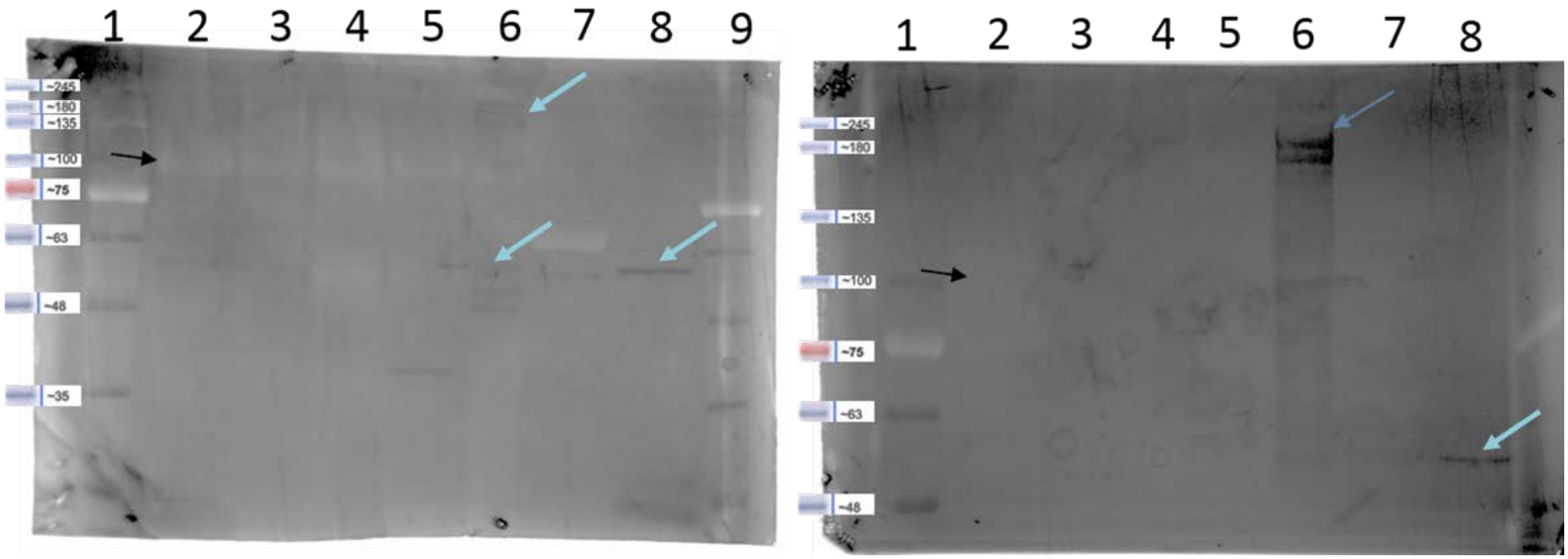


Figure 24. Western Blot analysis of the putative Avian HA RBD V3.4 protein from multiple purification steps using in ferret polyclonal Antisera (1:7500 dilution) (anti-H5N2, anti-H7N3 strains) to Avian Influenza A virus. Lane 1, Blueye Protein Ladder; Lane 2, HA3.4 PBS 7.2 (AEX, Size Exclusion purified), Lane3, #1 AEX Elution fraction 150 mM (Amicon >10Kda) retentate; Lane 4, #4 AEX Elution fraction 150+200 mM (Amicon >10Kda) retentate; Lane 5, #5 AEX Elution fraction 150 mM (Amicon >10Kda) retentate; Lane 6, HA 3.4 PBS 7.2 (First Mass-Spec Sample); Lane 7, BSA (3 ug); Lane 8 , Trivalent Sanofi Pasteur 2013 Vaccine; Lane 9, Blueye Ladder. In B) Western Blot analysis of HA3.4 protein from purification steps using ferret pAbs Antisera (1:7500 dilution) (H5N2, H7N3 strains) to Avian Influenza A virus. Lane B1, Blueye Protein Ladder; Lane B2, HA3.4 PBS 7.2 (AEX, Size Exclusion purify), LaneB3, #1 AEX Elution fraction 150 mM (Amicon >10Kda) retentate; Lane B4, #4 AEX Elution fraction 150+200 mM (Amicon >10Kda) retentate; Lane B5, #5 AEX Elution fraction 150 mM (Amicon >10Kda) retentate; LaneB 6, HA 3.4 PBS 7.2 (First Mass-Spec Sample); Lane B7, BSA (3 ug); Lane B8 , Trivalent Sanofi Pasteur 2013 Vaccine. Samples were run in a SDS-PAGE 4%-7% Semi-dry transferred to a PVDF membrane (25 V x 15 minutes). Secondary antibody Goat anti Ferret-HRP (1:10,000) as the detection reagent and Ultra-1-step-TMB solution as substrate. Black arrows indicate the position of the putative Avian HA RBD V3.4 protein in the membrane, blue arrows indicate reactive signals as in A2, A5, A6,A8, B6 and B8.

3.6. PCR mutagenesis and cloning of Influenza HA-RBD versions in the extracellular expression plasmid pPicZalpha-A and in the intracellular expression plasmid pPicZ-B, restore the open reading frame for c-Myc and 6-HIS tags.

Cloned PCR mutated versions were generated for the Avian HA-RBD mut-E-V3.4, and Avian HA-RBD mut-E-V3.6 versions (**Figure 25**),. Human HA-RBD mut-E-V3.3 was cloned in the extracelullar expression vector pPicZalpha-A only. The enzymatic ligation of each of the PCR XhoI /XbaI digested product was re-confirmed following restriction of the plasmid preparations obtained from transformed individual *E. coli* clone cultures with the same enzymes (**Figure 26, Figure 27**). Only Avian HA-RBD mut-E-V3.4 and Avian HA-RBD mut-E-V3.6 were fully sequenced and the results confirmed the correct removal of the premature Stop codon, and that the open-reading frame extended to include the C-Myc and 6-Histitdine tags contained in the pPicZalpha-A plasmid. The expected nucleotide sequences corresponding to the codon optimized multimeric avian genes were also confirmed (**Figure 28**). The consensus result represented the alignment of three plasmid preparation clones that were individually assembled from their sequencing results.

On the other hand, the PCR mutated version for only the Avian HA-RBD mut-I-V3.4, Avian HA-RBD mut-I-V3.6 and Human HA-RBD mut-I-V3.3 (**Figure 26, Figure 27)** were amplified, enzymatic digested and cloned in the intracellular (I) expression vector pPicZ-B. The incorporation fo the PCR XhoI /XbaI digested product into the plasmids were re-confirmed by applying enzymatic restriction of the plasmid preparations obtained from individual *E. coli* clones cultures with the same enzymes. Only pPicZB- Avian HA-RBD mut-I-V3.4 and pPicZB-Avian HA-RBD mut-I-V3.6 were

fully sequenced. The results confirmed the correct removal of the premature Stop codon at the 3’ end of the gene and that the open-reading frame extended to include the C-Myc and 6-Histitdine tags contained in the pPicZ-B plasmid. The expected nucleotide sequences corresponding tof the codon optimized multimeric avian genes and the correct addition of the Kozak consensus sequences for translation initiation were also confirmed (**Figure 29)**.

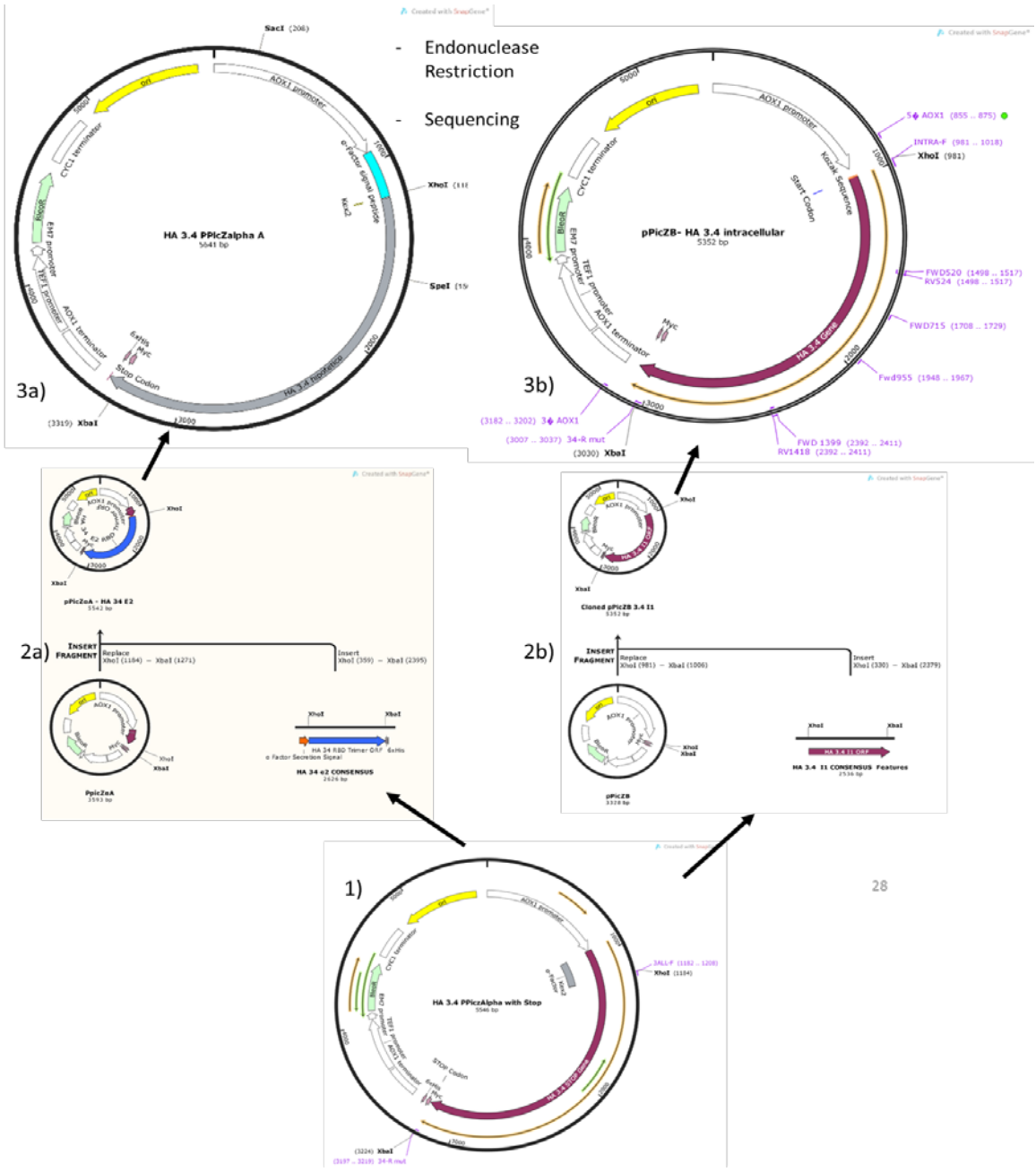


Figure 25. General cloning strategy for 6-His Tagged Avian and Human HA RBD mutated version. 1) A plasmid template pPicZalphaA-containing the Avian or Human HA-RBD version is used for PCR amplification. The PCR products and empty expression vector ((extracellular goes 2a), (intracellular goes 2b)) are restricted with specific endonucleases and their product ligated. In 3a) or 3b) Plasmid preps from transformed E coli cells are analyzed for the correct insertion then sequencing using specific primers to corroborate the mutated HA RBD versions. Plasmid from positives clones were amplified and used for transformation in *Pichia pastoris* X-33 electrocompetent cells.

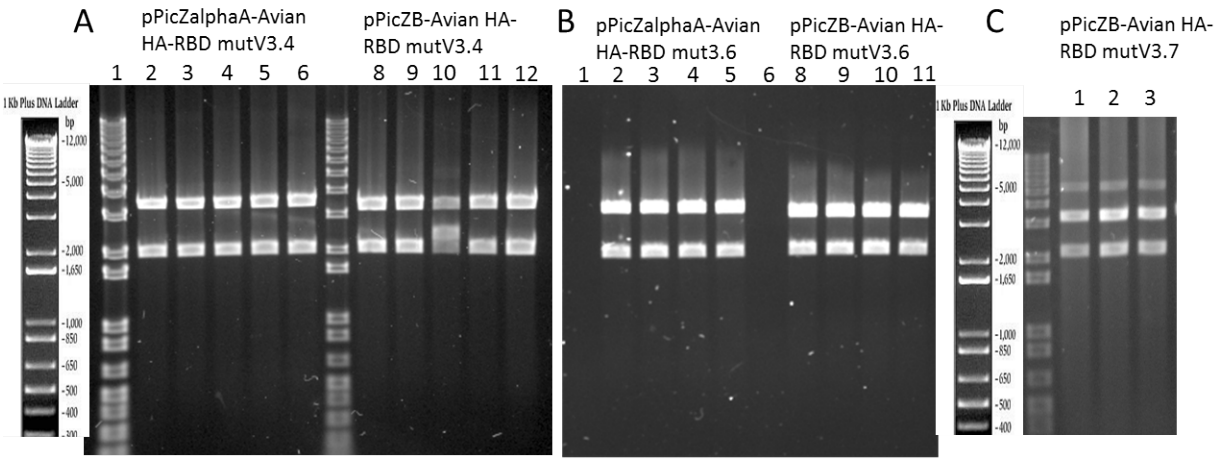


Figure 26. Restriction analysis of Avian HA-RBD mutV3.4 (A) and Avian HA-RBD mutV3.6 (B) and Avian HA-RBD mutV3.7 (C) using XhoI and XbaI endonucleases. In A, lanes (2-6) and lanes (8-12) show individual plasmid preparations (Top10 *E coli* clones) confirming that the genes were ligated into the plasmid backbone and transformed into E. coli. The ligation product of pPicZalpha and the Avian HA-RBD mut3.4 PCR XhoI/XbaI digested product and the product of pPicZB and the Avian HA-RBD mutHA3.4 PCR Xho/Xba digested product, were analyzed, respectively. In B, the same analysis was performed for the HA 3.6 (2.1 kB) gene (four clones, Lane 2-5) ligated into pPicZalphaA (3.3 kB) vector and the HA 3.6 (2.1Kb) gene (four clones, lanes 8 -11), ligated into the vector pPicZB (3.0 kb). All clones were positive for the HA RBD fragments. In C, restriction analysis was used to confirm the ligation of the Avian HA-RBD mutV3.7 into the intracellular expression plasmid pPicZB , one plasmid prep (Lane 1-3) from three clones was positive for the analysis.

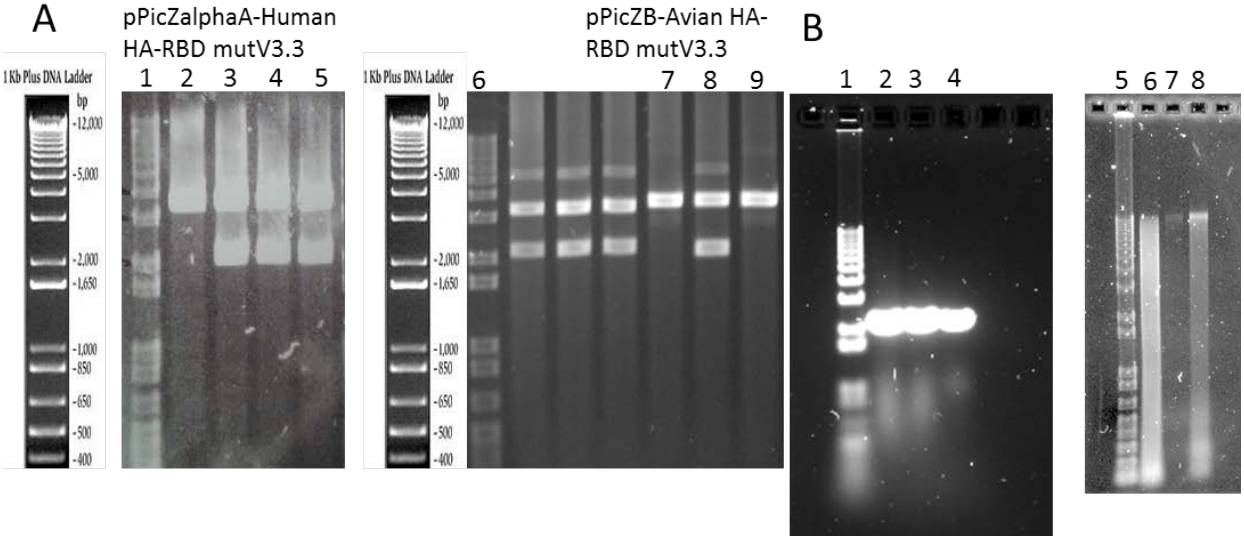


Figure 27. Restriction analysis of the gene Human HA-RBD mutV3.3 using XhoI and XbaI endonucleases (A), and its PCR amplification from the Human HA RBD V3.3 unmodified version integrated in *Pichia pastoris* genome (B). In A, Three (Lane 3-5) of four individual plasmids preps (Top10 E coli clones) confirming the gene ligation that were transformed with the ligation reaction of pPicZalpha + Avian HA-RBD mut3.3 PCR XhoI/XbaI digested product and in (lanes 7-9) one of three clones was positive for the ligation reaction pPicZB + Avian HA-RBD mutHA3.3 PCR Xho/Xba digested product, respectively. In B, Lane 2-4, PCR amplification (A) of the gene Human HA-RBD mutV3.3 (2.1 Kb), from the Pichia genome tranformed with the unmodified version of the gene (clone D7) for extracellular expression using specific primers and with two genomic DNA extraction methods applied, lanes 6-8. DNA Ladder standard are observed in lanes A1, A6, B1 and B5.

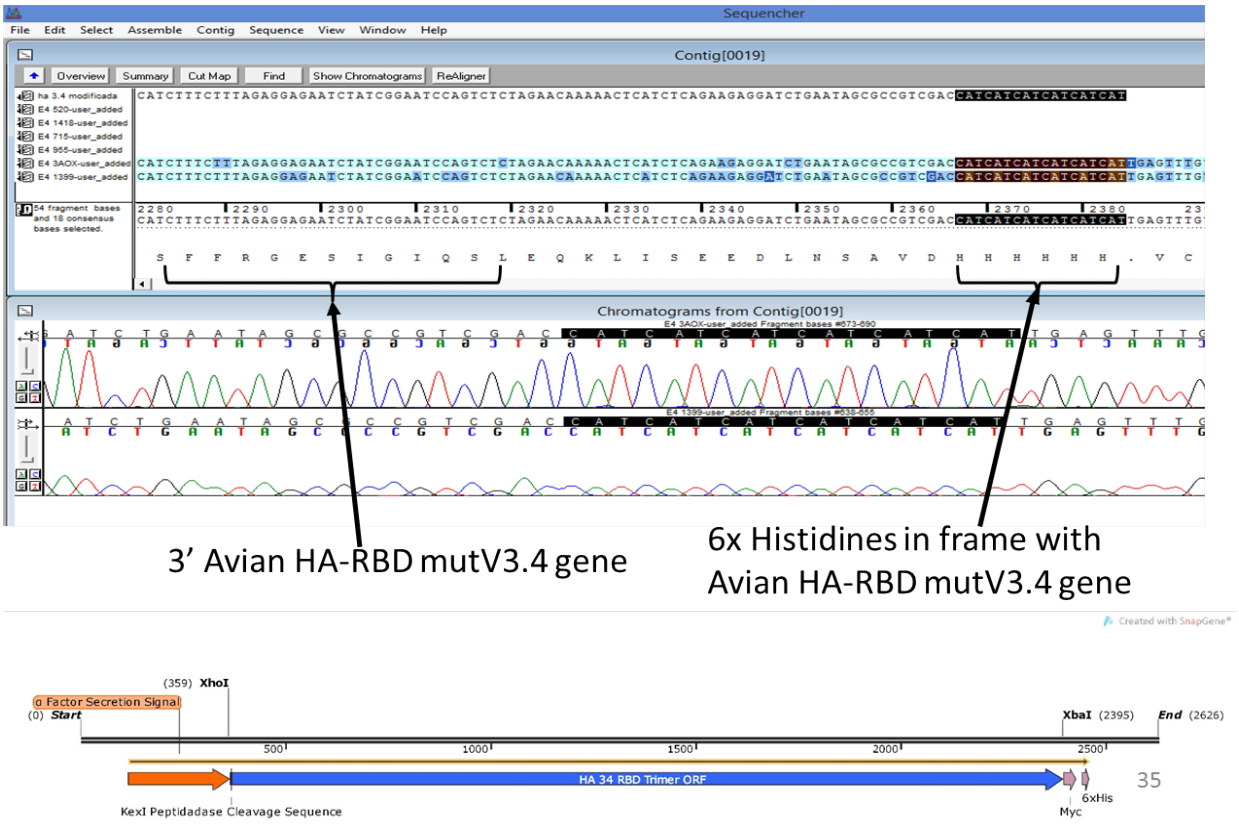


Figure 28. Sequencing result for the extracellular Avian HA-RBD mutV3.4 cloned in pPicZalphaA, showing the added features. There was no observed premature Stop codon which allowed the Avian HA-RBD mutV3.4 to be expressed in-frame with the c-myc and 6-His tags as indicated by the arrows which differed from the original Avian version. Identical sequencing results were obtained for the three different *E. coli* clones harboring the enzymatically confirmed multimeric gene. Below, there is a diagram showing the ORF features. Similar results applied to the Avian Extracellular and Intracellular HA-RBD mutV3.6 gene.

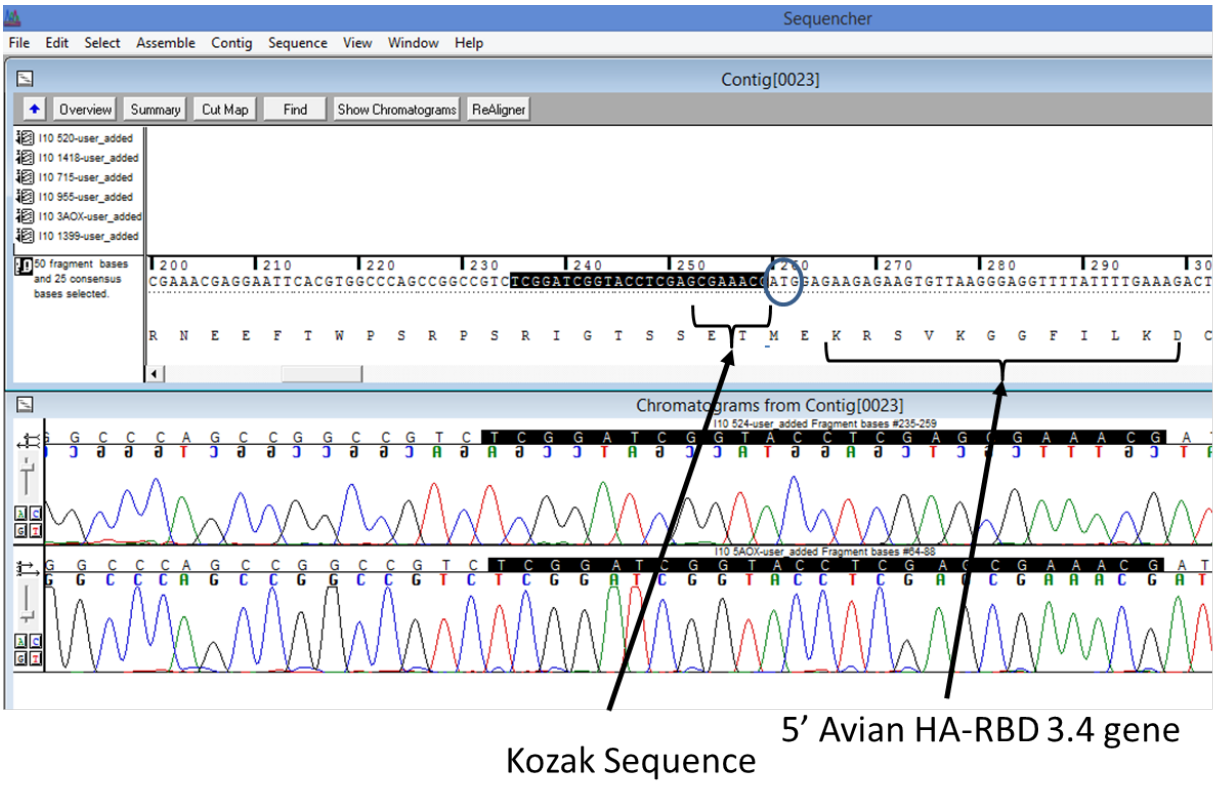


Figure 29. Sequencing result for the Avian HA-RBD mutV3.4 cloned into pPicZB, showing the added features. The sequence shows the Kozak consensus sequences, followed by the ATG start codon, and the 5' portion of the multimeric Avian gene. Sequencing results of the plasmids purified from three different *E. coli* clones harboring the enzymatic confirmed gene showed identical results. Similar results applied for the Avian Intracellular HA-RBD mutV3.6 gene.

3.7. *Pichia pastoris* transformed with the Human and Avian His-Tagged HA RBD extracellular versions, express highly and heterogenous N-glycosylated polypeptides recognized by ferret anti-sera to influenza, but lack of fuctional hemagglutination.

The mutated versions of the Avian and Human RBD generated by PCR and integrated into pPicZalphaA (Avian HA RBD mutV3.4, Avian HA RBD mutV3.6, Human HA RBD mutV3.3) and pPicZB (Avian HA RBD mutV3.4, Avian HA RBD mutV3.6 and Human HA RBD mutV3.3) were transformed in *Pichia pastoris* X-33. Zeocin hyperresistant clones for all the versions were selected. Methanol-induced protein expression by the Zeocin hyperresistance clones was performed for the extracellular secreted versions pPicZalphaA-Avian HA RBD mutV3.4, pPicZalphaA-Avian HA RBD mutV3.6, and pPicZalphaA-Human HA RBD mutV3.3; and for the pPicZB-Avian HA RBD mutV3.4 intracellular version. Only the three extracellular versions were able to generate detectable levels of protein expression in the supernatant of cultures treated with methanol for 48 h when recovered by Ni-NTA agarose chromatography (**Figure 30**). The recovered molecules were detected using a mouse anti-6Histidine-HRP monoclonal antibody in immunoblot analysis (**Figure 31**). The protein yield range from 140 µg/L for the Human HA RBD mutV3.3, 22µg/L for the Avian HA RBD mutV3.4, and 67.2 µg/L for the Avian HA RBD mutV3.6 transformed yeast (**Figure 30D**).

Coomasie blue staining of SDS-PAGE, under reducing conditions (Figure 29), and anti-Histidine monoclonal antibody immunoblot analysis for the two Avian versions, HA RBD mutV3.4, and HA RBD mutV3.6, detected a high molecular weight protein band between 135-235KDa. These observed bands differ from the expected (~79 KDa) calculated sequences. Furthermore, enzymatic deglycosylation assays performed on

these two avian version, demonstrate that the recovered proteins are highly N-glycosylated. Deglycosylation produces de-glycosylated protein bands in the range of 77KDa and 16KDa as detected by Anti-His antibody. Based on extrapolation of their relative mobility values (**Figure 32, Table 19**), structural models were generated based on the de-glycosylated results (**Figure 34**) that suggest a full-size glycosylated Avian HA RBD mutV3.4 and Avian HA RBD mutV3.6 (V3.4 core)(**Figure 35**)

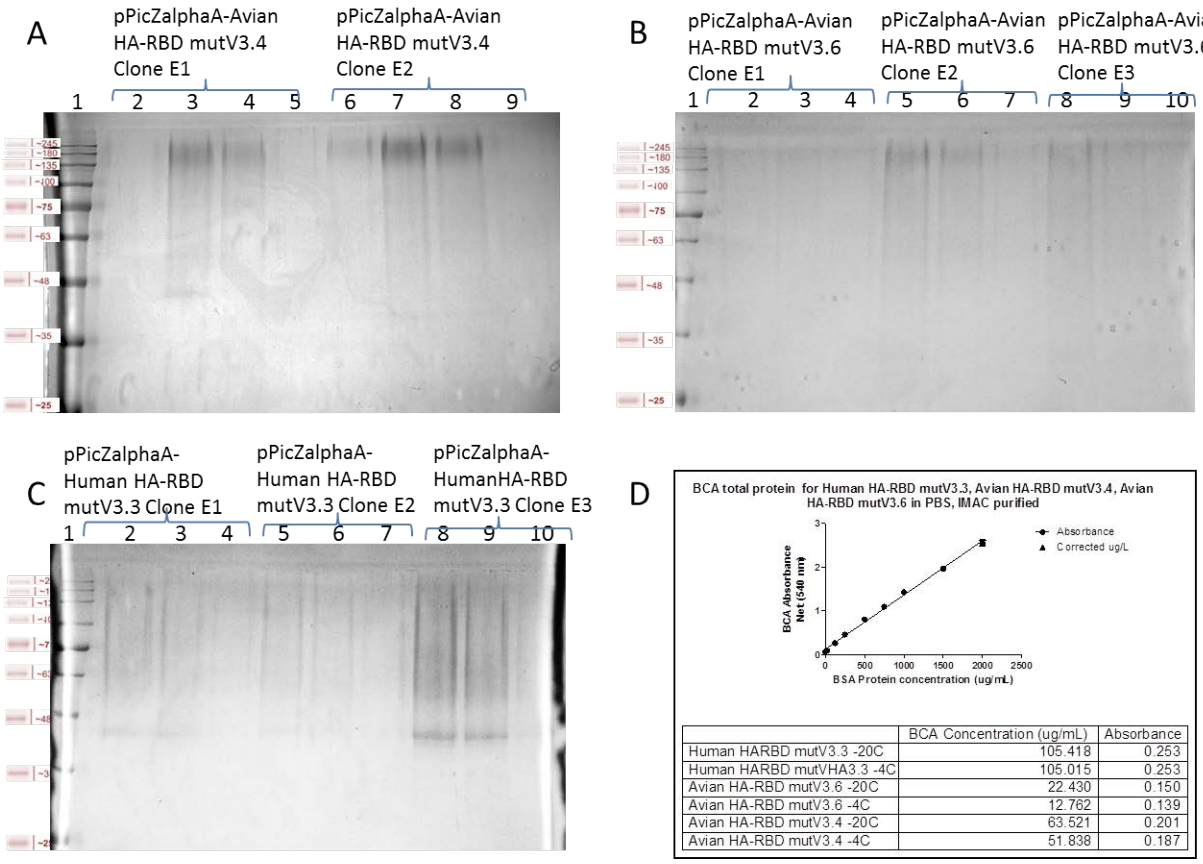
In the case of the Human HA RBD mutV3.3 immunoblot analysis using Anti-6 histidine monoclonal antibodies, detected a protein smear on SDS-PAGE analysis (**Figure 31**). Enzymatic deglycosylation using the EndoHf enzyme followed by Anti-His monoclonal immonoblot detection for this protein, produced mostly a low molecular ~35 KDa peptide (**Figure 231**) (**Table 19**) suggesting a heterogeneous N-glycosylated and mostly truncated Human HA RBD V.3.3 protein form (**Figure 36**) based on cleavage site analysis.

Contrary to the extracellular strategy, no protein expression was detected for the intracellular Avian HA RDB mutV3.4 strategy tested to date.

Antibody characterization using ferret antisera to the influenza virus resulted in the detection of the Avian HA-RBD mutV3.4 by the ferret immunized antisera: H5 subtyping (FR-229), H5 subtyping (FR-231:ferret anti-sera to Influenza A/Duck/Hunan/795/2002(H5N1)) and H7 subtyping **(**FR-1088: ferret anti-sera to Influenza A/Turkey/Virginia/2002 (H7N2)/PR8/BCDC-5 (**Figure 33A**). Avian HA-RBD mutV3.6 samples was recognized by H5 subtyping (FR-231:ferret anti-sera to influenza A/Duck/Hunan/795/2002(H5N1, slightly recognized by H7 subtyping **(**FR-1088 :ferret

anti-sera to Influenza A/Turkey/Virginia/2002 (H7N2) /PR8/BCDC-5) (**Figure 33B**). The Human HA-RBD mutV3.3 was recognized by influenza B subtyping FR-1080 (ferret anti-sera to B influenza virus B/Texas/06/2011) but was not recognized by H1 subtyping (FR-359: ferret anti-sera to influenza A virus A/Cal/07/2009 H1N1 pdm09) or H3 subtyping (FR-1000 (WHO Supplemental Antiserum, influenza A(H3N2)v reference ferret antiserum) (**Figure 33C)** adding evidence to protein truncation and only detection to the B type subunit.

Hemmaglutination assays fail to detect functional activity for the HA-RBD mutV3.3, HA-RBD mutV3.4, HA-RBD mutV3.6 proteins, but were positive to hemmaglutination, the samples Trivalent Vaccine Fluviral (2015-2016), Hemagglutnin Inhibition HA positive control, used to validate the assay.



	BCA Concentration (ug/mL)	Absorbance
Human HARBD mutV3.3 -20C	105.418	0.253
Human HARBD mutVHA3.3 -4C	105.015	0.253
Avian HA-RBD mutV3.6 -20C	22.430	0.150
Avian HA-RBD mutV3.6 -4C	12.762	0.139
Avian HA-RBD mutV3.4 -20C	63.521	0.201
Avian HA-RBD mutV3.4 -4C	51.838	0.187

Figure 30. SDS-PAGE analysis for expression of the Avian and Human HA-RBD versions. After 48 h of methanol induction for protein expression under the control 0f the promoter AOX1, cultures were harvested and purified using Ni-NTA agarose beads. In A, in lanes 2-5, elution profile (first to fourth in Elution Buffer (250 mM imidazole in phosphate Buffer, pH 8.0) of clone E1 for Avian HA-RBD mutV3.4; lanes 6-9 elution profile (first to fourth elution in Elution Buffer) of clone E2 for the Avian HA-RBD mutV3.4. In B, in lanes 2-4, elution profile (first to third elution in Elution Buffer) of clone E1 for the Avian HA-RBD mutV3.6; lanes 5-7 elution profile (first to third elution in Elution Buffer) of clone E2 for the Avian HA-RBD mutV3.6; lanes 8-10, elution profile (first to third elution in Elution Buffer) of clone E3 for the Avian HA-RBD mutV3.6. In C, Human Versions HA RBD mutV3.3 for three clones E1 (lane2-4), E2 (lane 5-7), E3 (lane 8-10), with their three elution samples, respectively. Clones E2 from Avian HA-RBD mutV3.4, clone E2 from Avian HA-RBD mutV3.6, and Clone E3 from Human HA-RBD mutV3.3, were selected for further protein analysis. In D, RBD-HA mutV3.3, RBD-HA mutV3.6, RBD-HA mutV3.7 quantified after Ni-NTA protein recovery and determined by the BCA method for after IMAC purification, ultrafiltration, desalted and corrected for final volumen in PBS, pH 7.4.

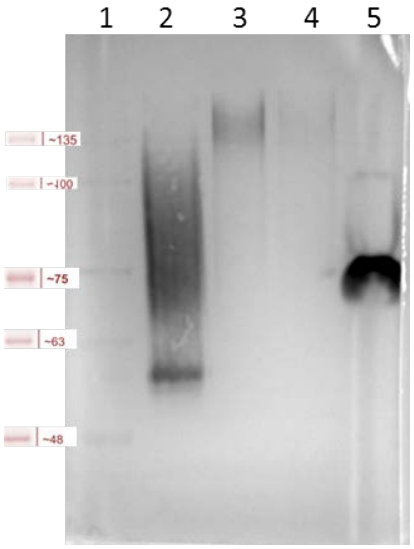


Figure 31. Immunodetection of the selected clones for the Avian and Human His-Tag HA-RBD mutVersions. Samples were purified by IMAC and immune-detected with mouse Anti-6His-HRP monoclonal antibody. In lane 2, Human HA RBD mutV3.3 produced a highly heterogeneous molecular size smear, in lane 3, Avian HA RBD mutV3.4 and lane 4,Avian HA RBD mutV3.6 are immunodetected as high molecular size recovered proteins between 135 KDa and 245 KDa. In lane 5, Vhh nanobody as a positive control for 6-Histidine tag detection.

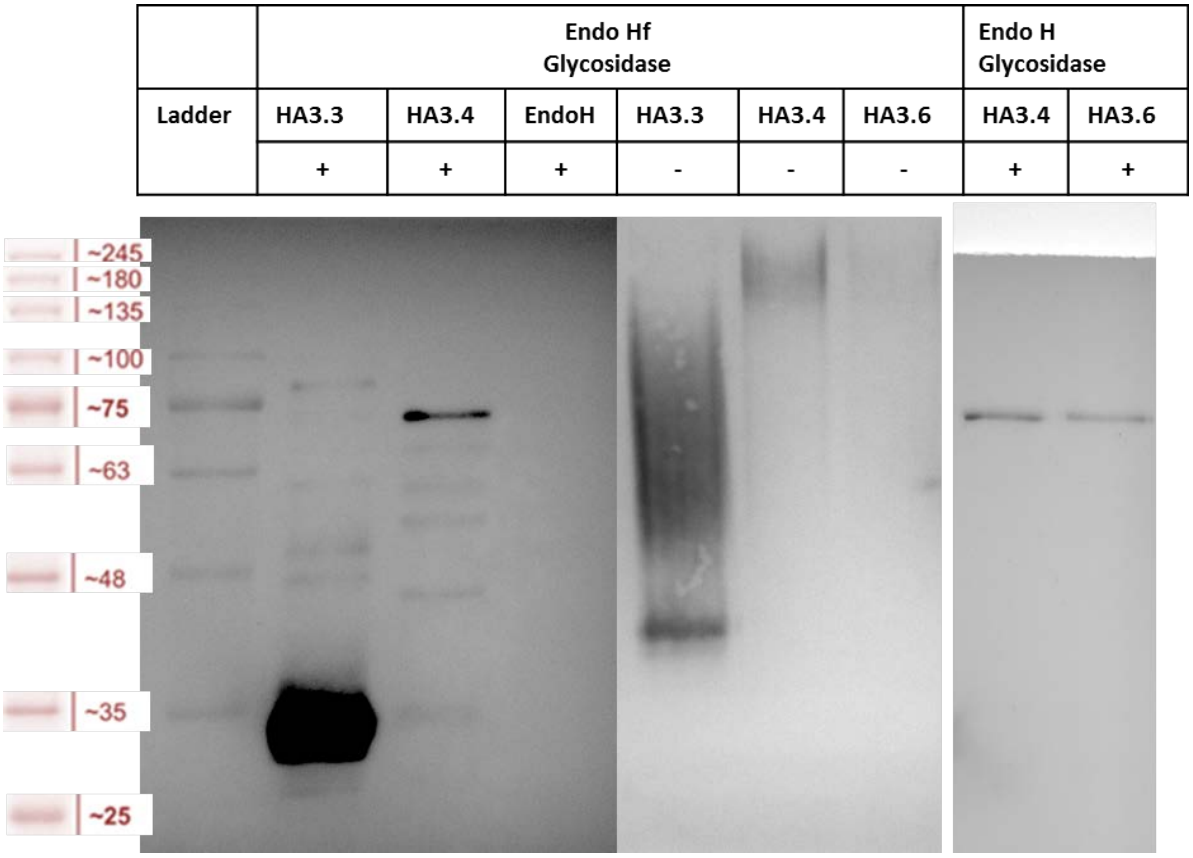


Figure 32. Deglycosylation enzymatic assays on Avian and Human RBD mutVersions. Avian HA-RBD mutV3.4 and Avian HA-RBD mutV3.6 were treated with EndoHf or EndoH glycosidase generating a discrete and sharp band with a molecular size ~75 KDa, which is close to the expected ~80 KDa molecular weight of their protein sequences, thus indicating a highly N-Glycosylated Avian glycoproteins (HA3.4(-) and HA3.6(-)). The Human HA RBD mutV3.3 is highly N-glycosylated, with heterogeneous glycoforms reduced to mostly a single discrete, sharp, and truncated ~35KDa molecular weight band after Endoglycosidase treatment, as calculated in Table 18.

Table 21. Molecular weight determination by SDS-PAGE for the Avian and Human Glycosylated and Deglycosylated forms. In A, Blueye protein standard for Molecular weight determination from Relative mobility (Rf). In B, Calculated Molecular weight for the discrete band observed in the Figure 30. (*) Indicates the most conspicuous immunodetected band after deglycosylation treatment.

A

Total distance (mm)	54		
Protein Standard, Blueye Ladder (KDa)		log MW	Rf
245	2	2.3892	0.03704
180	3.5	2.2553	0.06481
135	5.5	2.1303	0.10185
100	9	2	0.16667
75	13	1.8751	0.24074
63	17.5	1.7993	0.32407
48	23.5	1.6812	0.43519
35	32.5	1.5441	0.60185
25	42.5	1.3979	0.78704
20	47	1.301	0.87037
17	51.5	1.2304	0.9537

B

	D (mm)	Rf	y Linear Regression	10(Y)=MW	Y linear regression	M.W (KDa)
HA3.3 endoHf	11.5	0.21296	2.012191	102.8468513	1.943376	87.776043
HA3.3 endoHf	22	0.40741	1.792629	62.03388777	1.751502	56.428954
HA3.3 endoHf	33.5	0.62037	1.552155	35.65783738	1.541353	**34.781876***
HA3.4 endoHf	15	0.27778	1.939003	86.89664319	1.879418	**75.756168***
HA3.4 endoHf	22	0.40741	1.792629	62.03388777	1.751502	56.428954
HA3.4 endoHf	26.5	0.49074	1.69853	49.94936828	1.669269	46.694852
HA3.4 endoHf	34.5	0.63889	1.531244	33.98161381	1.523079	33.348707
HA3.4 endoHf	35	0.64815	1.520789	33.17332473	1.513942	32.654422
HA3.4 endoHf	38.5	0.71296	1.447601	28.02857381	1.449983	28.182726
HA3.4 endoHf	51	0.94444	1.186217	15.35383963	1.221561	16.655628
HA3.6 endoHf	14.5	0.26852	1.949459	89.01413994	1.888554	**77.366687***
HA3.6 endoHf	21.5	0.39815	1.803084	63.54538276	1.760638	57.628591
HA3.6 endoHf	26	0.48148	1.708986	51.16653411	1.678407	47.687768
HA3.6 endoHf	34	0.62963	1.541699	34.80959737	1.532215	34.057675
HA3.6 endoHf	34.5	0.63889	1.531244	33.98161381	1.523079	33.348707
HA3.6 endoHf	38	0.7037	1.458056	28.71150778	1.45912	28.781936
HA3.6 endoHf	50.5	0.93519	1.196672	15.72794566	1.230698	17.009753
HA3.4 Negative EndoHf	4	0.07407	2.169022	147.578129	2.08043	**120.34554***

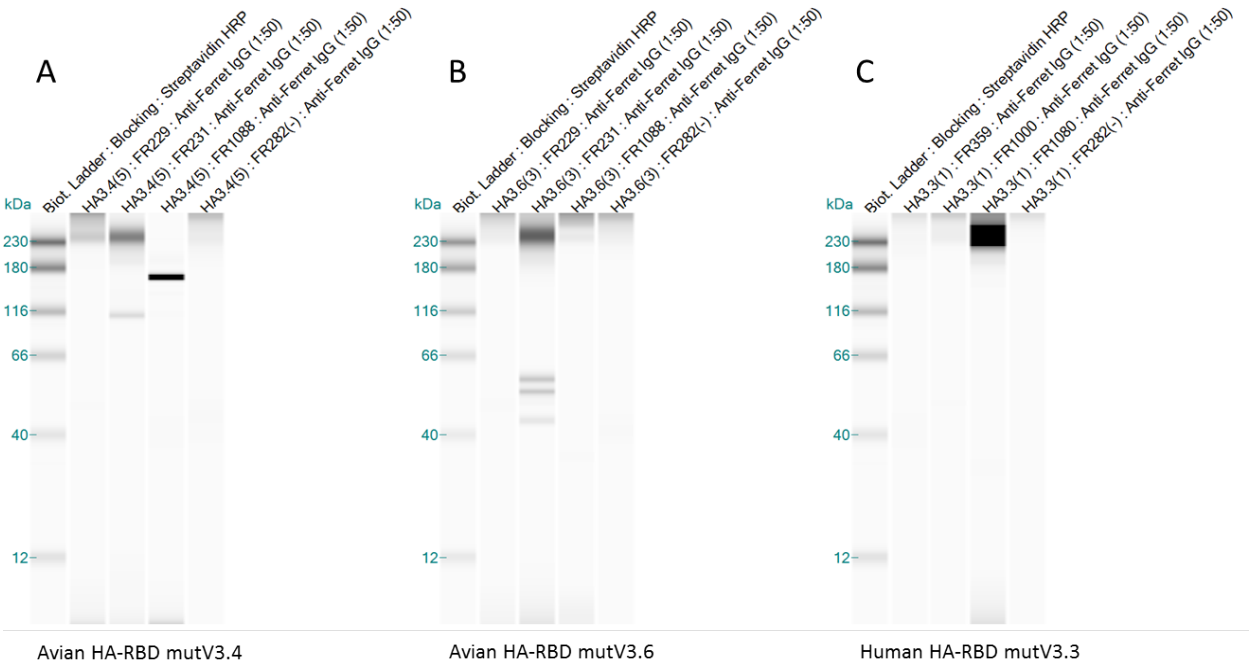


Figure 33. WES analysis for ferret anti-influenza polyclonal characterization of the glycosylated Avian and Human HA-RBD multimeric genes, purified by IMAC. In A, Avian HA-RBD mutV3.4 transformed yeast expressed protein with high molecular sizes of ~230 KDa which were weakly recognized by the antisera directed against FR-229 (H5 subtype). The FR-231 (Ferret anti-sera to Influenza A/Duck/Hunan/795/2002 (H5N1)) antisera recognized two bands of 230KDa and ~116KDa. The FR-1088 (Ferret anti-sera to Influenza A/Turkey/Virginia/2002 (H7N2)/PR8/BCDC-5) reacted strongly with a 165KDa band. In B, Avian HA-RBD mutV3.6 samples were not detected by the FR-229 (H5 subtype) antisera, but were recognized by the FR-231 (ferret anti-sera to influenza A/Duck/Hunan/795/2002(H5N1)) antisera, with one high molecular size band at 230 KDa, and were weakly recognized by the FR-1088 (ferret anti-sera to influenza A/Turkey/Virginia/2002 (H7N2) /PR8/BCDC-5) antisera. In C, the Human HA-RBD mutV3.3 expressed proteins were recognized by FR-1080 (ferret anti-sera to B Influenza virus B/Texas/06/2011) antisera with a molecular size of 239 KDa, but was not recognized by the FR-359 (ferret anti-sera to Influenza A virus A/Cal/07/2009 H1N1 pdm09) or FR-1000 (WHO Supplemental Antiserum, Influenza A(H3N2)v Reference Ferret Antiserum) antisera. All samples were compared to a non-immune FR-282 (ferret negative Control sera) antiserum in A, B and C analysis.

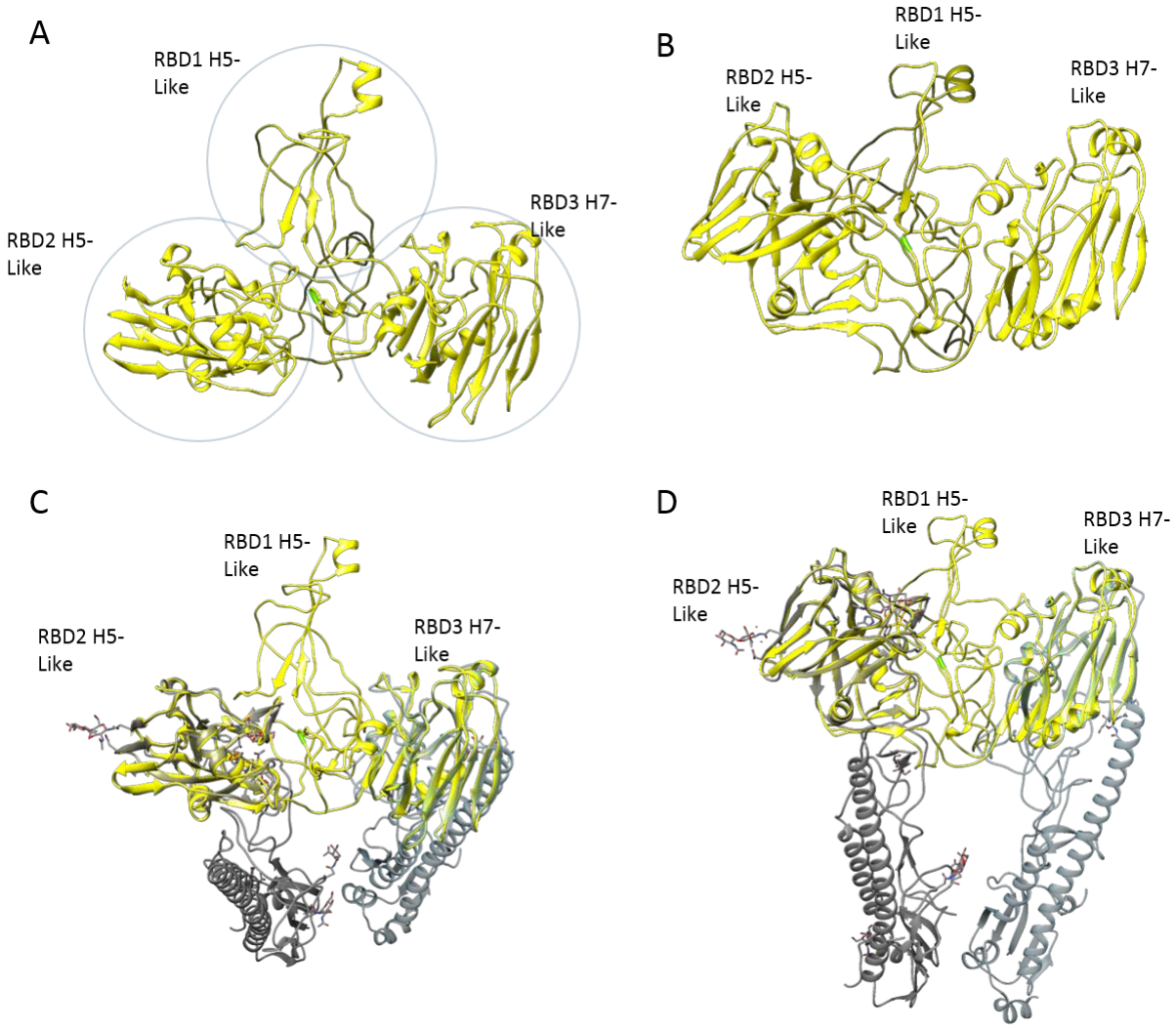
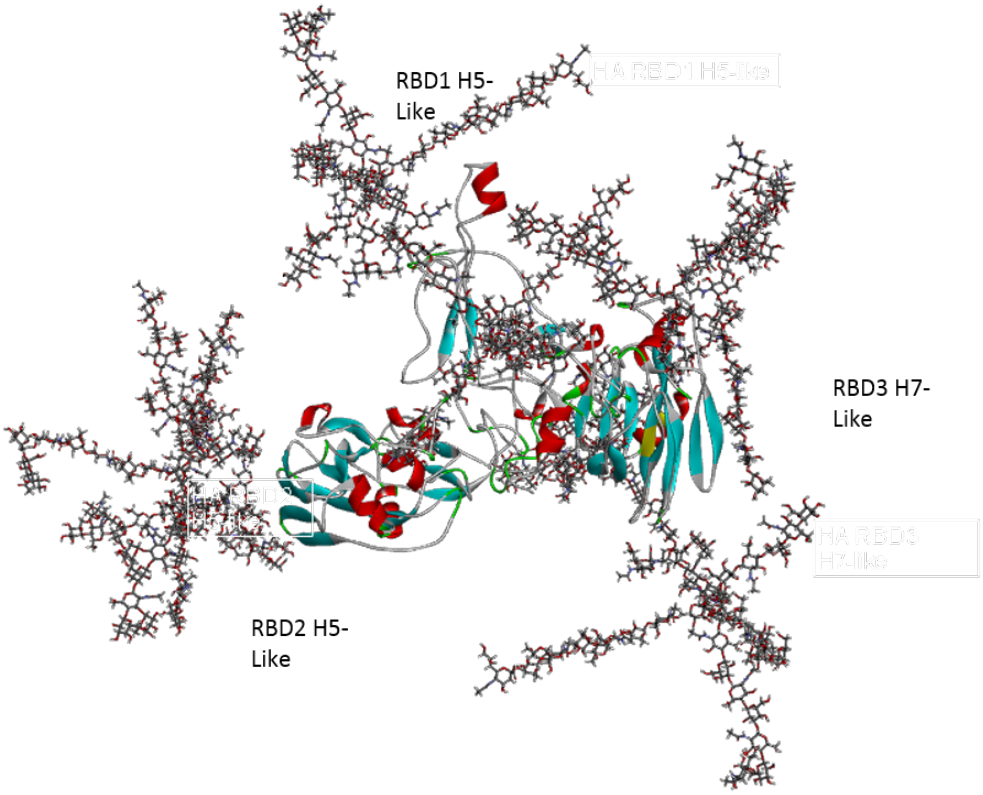


Figure 34. Structural model of the Avian HA-RBD mutV3.4. The PDB file for this multimeric protein sequence was generated by the Phyre2 server (72). RBD1 subunit fail to generate a folded structure for the whole multimeric model, therefore, the two remaining regions RBD2 and RBD3 were superimposed to the determined crystal structure maps of the Hemagglutinin H5 (in green) of the virus A/duck/Egypt/10185SS/2010 (H5N1) (PDB file 5E2Y) (79) and the Hemmaglutinin H7 of the virus A/Anhui/1/2013 (PDB file 4R8W) (80). The graphs were generated using Chimera UCSF software, crystal structure PDB file retrieved from the RCSB Protein Data Bank website. In A, Avian HA RBD V2.4 (top view). In B, same as A (lateral view). In C (top view) and D (lateral view), Avian HA-RBD V3.4 model superimposed to crystal structure 5E2Y(H5 hemagglutinin) and 4R8W (H7 hemagglutinin).



Avian HA RBD mutV3.4 Sequence length: **675 AA.**

8 potential N-glycosylation sites are found.

Avian HA RBD mutV3.4 Sequence length: **675 AA.**

8 potential N-glycosylation sites are found.

```
  1: SVKGGFILKDCSVAGWLLGN PMCDEFLTAPEWSYIVEKDN PSNDLCYPGNFDNYEELKHL MSSTNHMERIQIFPRNSWTN
 81: HNASAGVSSACPFNGGSSFF RNVVWLIKKNDVYRTLKRNY TNTNVEDLLIIWGVHHPNDA TEQTKLYQNLNTYVSVGTST
161: LNQRSTPTIATRPQVNGQRG RMEFFWTILKSNDSIFFEST GNFIAPEYAYKIVKKGKSTV MKSGGGGSGGGGSPLILKDC
241: SVAGWLLGNPMCDEFLNVQE WSYIVEKDNPSNDLCYPGNF DNYEELKHLMSSTNHLERIQ IFPRNSWPNHDSSSGVSSAC
321: PFNGESSFFRNVVWLIKKND VYRTLKRNYTNTNVEDLLII WGVHHPNDATVQTKLYQNLN TYVSVGTSTLNQRSIPTIAT
401: RPQVNGQRGRVEFFWTILES NDSIFFESTGNFIAPEYAYK VTKKGKSAVMKSGGGGSGGG GSPTDLGQCGLLGTLIGPPQ
481: CDQFLKFDADLIIERREGTD VCYPGKFTNEESLRQILRRS GGIDKESMGFTYSGIRTNGT TSACRRSNPSFYAEMKWLLS
561: NSDNATFPQMTKSYRNPRNK PALITWGVHHSGSATEQTKL YGSGDKLITVGSSKYLQSFT PSPGARPQVNGQSGRIDFHW
641: LLLDPNDTVTFTFNGAFIAP DRASFFRGESIGIQS
```

Stage 1e - 7 potential N-glycosylation sites are accessible

No.	AA Position	PDB Residue	Chain Position	Torsion Angle [0-360] N_CA_CB_CG|CA_CB_CG_OD1|C1_ND2_CG_CB|O5_C1_ND2_CG|			
1	82	82	82	60 °	340 °	160 °	260 °
2	192	192	192	220 °	40 °	160 °	260 °
3	348	348	348	60 °	340 °	160 °	260 °
4	421	421	421	180 °	340 °	160 °	260 °
5	538	538	538	180 °	340 °	160 °	260 °
6	564	564	564	180 °	340 °	160 °	260 °
7	646	646	646	180 °	340 °	160 °	260 °

Figure 35. In silico N-glycosylation modelling for the Avian HA-RBD mutV3.4. The Avian HA-RBD mutV3.4 PDF file obtained from Phyre2 server (72) were modelled in the GlyProt server to calculate the geometric permitted N-Glycosylation sites. A total of 7 potential and geometrically permitted glycosylation motifs were detected and in-silico constructed under a “very large carbohydrate branch” assumption.

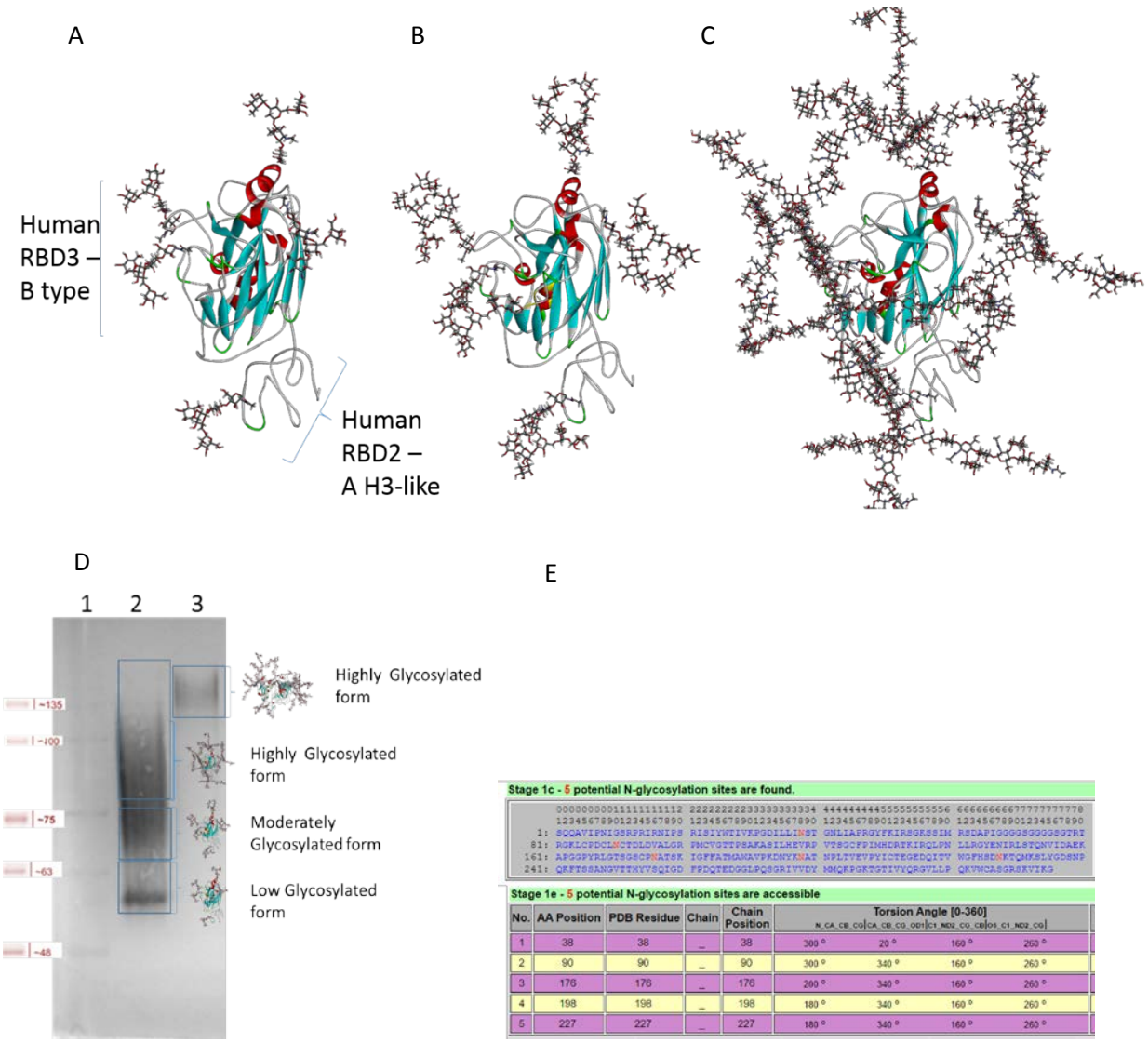


Figure 36. In silico N-glycosylation modelling for the truncated Human HA-RBD mutV3.3 showing different levels of N-glycosylation synthesis and its effect on protein heterogeneity. N-Glycosylation profile were generated with GlyProt Server, using the PDB files generated by Phyre2 server (72). In A, basic N-Glycosylation modelling profile showing the complete RBD3 subunit comprising the Influenza B RBD region, and also showing the human RBD2 subunit partially proteolysis. In B, low mannose N-glycosylation model, corresponding a moderate glycosylated form as seen in sequence in D. In C, a very large N-glycosylated protein form that correspond to its highly glycosylated protein in figure D. In D, protein heterogeinty of N-glycosylated forms, in lane 1, blueye protein ladder; lane 2, Human HA RBD mutV3.3, showing discrete forms of glycosylations; lane 3, Avian HA RBD mutV3.4 showing its higly glycosylated form, corresponding to the hypothetical glcyo-form in C. In E, Avian HA-RBD sequences used for the structural model, and the five N-glycosylation sites geometrically permitted for the in-silico modelling.

Dilution →	1:1	1:2	1:4	1:8	1:16	1:32	1:64	1:128	1:256
HA RBD mutV3.3 [Centricon Concentrated]	-	-	-	-	-	-	-	-	-
HA RBD mutV3.4 [Centricon Concentrated]	-	-	-	-	-	-	-	-	-
HA RBD mutV3.3 (5.25 ug)	-	-	-	-	-	-	-	-	-
HA RBD mutV3.4 (3.15 ug)	-	-	-	-	-	-	-	-	-
HA RBD mutV3.6 (1.1 ug)	-	-	-	-	-	-	-	-	-
Flurival 2014-2015 Influenza Vaccine [Concentrated]	+	+	+	+	+	+	+	+	+
Influenza A virus (H1N1) control antigen, BPL inactivated (FR-187)	+	+	+	+	+	+	-	-	-
PBS	-	-	-	-	-	-	-	-	-
Flurival 2014-2015 Influenza Vaccine	+	+	+	+	+	+	+	+	+

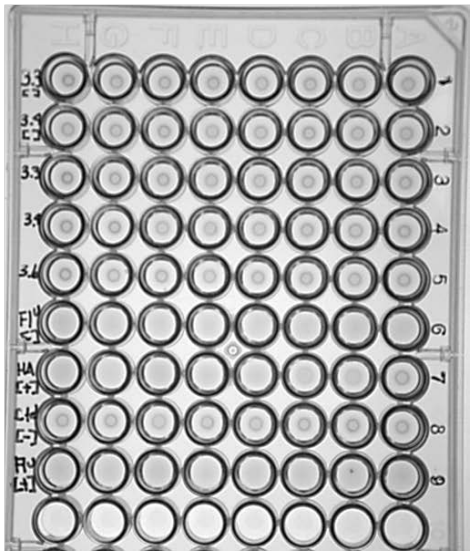


Figure 37. Hemagglination assays in RBC guinea pig erythrocytes with the HA-RBD recombinant proteins. Red blood cells where incubated for 30 min. with serial dilutions of HA RBD proteins and their hemagglutination were evaluated qualitatively, observing a lack of functional (round botton in the well) hemagglutination propierties. Positive control (diffused cells across the well forming a homogenous sieve) comprising the Inlfuenza vaccine Fluviral and Influenza A control antigen FR-187, were used to compare results.

3.8. Human HA-RBD mutV3.3 peptide truncation is explained by a predicted internal pro-peptide cleavage site.

One internal protease cleavage signal for Human HA RBD mutV3.3 is predicted by ProP v1.0 software (**Table 21)**, and is expected to result in a peptide with a similar calculated molecular weight of ~36KDa (**Table 22**) when compared to the observed deglycosylated peptide with a molecular weight of ~35 Kda. On the other hand, no predicted internal protease signal is observed within the peptide sequence of the Avian HA RBD mutated V3.7 (**Table 23**).

The effect of amino acid substitution within the predicted sequence was determined using ProP v1.0 (**Table 24**). Substitutions present in nine of twelve sequences scored numbers below the predicted threshold value of the wildtype sequence. Four of these nine sequences, Mut6 (K/A), Mut7 (R/A), Mut10 (K/R), Mut12 (K/R)/(R/K) scored the lowest values with the dibasic Kex2 cleavage site (KR) involved. When alanine-scanning mutagenesis is used, the lowest values correspond to Mut6 (K/A)(**0.063**) and Mut7 (R/A)(**0.059**). Alternatively, basic substitution in Mut10 (K/R)(**0.140**), or the double basic mutant Mut12 (K/R)/(R/K)(**0.100**) are predicted to have an negative effect in cleavage site recognition.

Table 22. Protease-cleavage site predicted for the Human HA-RBD mutV3.3. The peptide sequence was analyzed using the propeptide cleavage site software tool ProP 1.0 server. Two predicted sites were detected above the threshold, the first (in green) correspond to to the signal peptide (Kex2) sequence endogenous to the post-translational processing signals designed in the *Pichia pastoris* expression system, and the second site (highlighted in red) corresponded to an internal signal peptide at the position (410) above the threshold value of the amino acid input sequence that is consistent with the experimental result for the de-glycosylation assays observed for the truncated HA RBD mutV3.3 (MW ~35KDa) peptide (Figure 32).

```
ProP v.1.0b ProPeptide Cleavage Site Prediction
General PC cleavage site prediction (Arginine/Lysine residues)

  Human HA RBD mutV3.3 Sequence
TIASIAAKEEGVSLEKRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETPSDNGTCYPGDFIDYEELREQLSSVS      80
SFERFEIFPKTSSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLSKSYINDKGKEVLVLWGIHHPSTSADQQ     160
SLYQNADAYVFVGSSRYSKKFKPEIAIRPKVRDQEGRMNYYWTLVEPGDKITFEATGNLVVPRYAFAMERNAGSGIIISD     240
GGGGSGGGGSGEICDSPHQILDGENCTLIDALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTL     320
EFNNESFNWTGVTQNGTSSACIRRSNNSFFSRLNWLHRLNFKYPALNVTMPNNEQFDKLYIWGVHHPGTDKEQIFLYAQS     400
SGRITVSTKRSQQAVIPNIGSRPRIRNIPSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIMRSDAPIGGGG     480
SGGGGSGTRTRGKLCPDCLNCTDLDVALGRPMCVGTTPSAKASILHEVRPVTSGCFPIMHDRTKIRQLPNLLRGYENIRL     560
STQNVIDAEKAPGGPYRLGTSGSCPNATSKIGFFATMAWAVPKDNYKNATNPLTVEVPYICTEGEDQITVWGFHSDNKTQ     640
MKSLYGDSNPQKFTSSANGVTTHYVSQIGDFPDQTEDGGLPQSGRIVVDYMMQKPGKTGTIVYQRGVLLPQKVWCASGRS     720
KVIKGLEQKLISEEDLNHHHHHH                                                              800
................P.............................................................      80
................................................................................     160
................................................................................     240
................................................................................     320
................................................................................     400
.........P......................................................................     480
................................................................................     560
................................................................................     640
................................................................................     720
........................                                                             800

Propeptide cleavage sites predicted:  Arg(R)/Lys(K): 2
```

Name	Pos	Context v	Score	Pred
Sequence	8	IASIAAK|EE	0.071	.
Sequence	16	EGVSLEK|RG	0.087	.
Sequence	17	GVSLEKR|GV	0.952	*ProP*
Sequence	26	APLHLGK|CN	0.074	.
Sequence	73	IDYEELR|EQ	0.044	.
Sequence	84	SVSSFER|FE	0.215	.
Sequence	90	RFEIFPK|TS	0.097	.
Sequence	101	PNHDSNK|GV	0.092	.
Sequence	113	CPHAGAK|SF	0.080	.
Sequence	117	GAKSFYK|NL	0.093	.
Sequence	124	NLIWLVK|KG	0.075	.
Sequence	125	LIWLVKK|GN	0.138	.
Sequence	131	KGNSYPK|LS	0.078	.
Sequence	134	SYPKLSK|SY	0.066	.
Sequence	140	KSYINDK|GK	0.058	.
Sequence	142	YINDKGK|EV	0.064	.
Sequence	176	VFVGSSR|YS	0.074	.
Sequence	179	GSSRYSK|KF	0.048	.
Sequence	180	SSRYSKK|FK	0.271	.
Sequence	182	RYSKKFK|PE	0.061	.
Sequence	188	KPEIAIR|PK	0.034	.
Sequence	190	EIAIRPK|VR	0.084	.
Sequence	192	AIRPKVR|DQ	0.125	.
Sequence	197	VRDQEGR|MN	0.060	.
Sequence	210	LVEPGDK|IT	0.065	.
Sequence	223	GNLVVPR|YA	0.048	.
Sequence	230	YAFAMER|NA	0.083	.
Sequence	284	CDGFQNK|KW	0.071	.
Sequence	285	DGFQNKK|WD	0.048	.
Sequence	292	WDLFVER|SK	0.199	.
Sequence	294	LFVERSK|AY	0.101	.
Sequence	311	PDYASLR|SL	0.033	.
Sequence	343	TSSACIR|RS	0.036	.
Sequence	344	SSACIRR|SN	0.370	.
Sequence	352	NNSFFSR|LN	0.040	.
Sequence	358	RLNWLHR|LN	0.048	.
Sequence	362	LHRLNFK|YP	0.084	.
Sequence	378	NNEQFDK|LY	0.071	.
Sequence	391	HHPGTDK|EQ	0.089	.
Sequence	403	YAQSSGR|IT	0.278	.
Sequence	409	RITVSTK|RS	0.066	.
Sequence	410	ITVSTKR|SQ	0.527	*ProP*
Sequence	422	IPNIGSR|PR	0.035	.
Sequence	424	NIGSRPR|IR	0.066	.
Sequence	426	GSRPRIR|NI	0.262	.
Sequence	431	IRNIPSR|IS	0.053	.
Sequence	440	IYWTIVK|PG	0.064	.
Sequence	457	GNLIAPR|GY	0.049	.
Sequence	461	APRGYFK|IR	0.069	.
Sequence	463	RGYFKIR|SG	0.031	.
Sequence	466	FKIRSGK|SS	0.049	.
Sequence	471	GKSSIMR|SD	0.110	.
Sequence	489	GGGSGTR|TR	0.269	.
Sequence	491	GSGTRTR|GK	0.176	.
Sequence	493	GTRTRGK|LC	0.080	.
Sequence	510	LDVALGR|PM	0.033	.
Sequence	521	GTTPSAK|AS	0.055	.
Sequence	529	SILHEVR|PV	0.035	.
Sequence	542	FPIMHDR|TK	0.032	.
Sequence	544	IMHDRTK|IR	0.100	.
Sequence	546	HDRTKIR|QL	0.040	.
Sequence	553	QLPNLLR|GY	0.041	.
Sequence	559	RGYENIR|LS	0.041	.
Sequence	570	NVIDAEK|AP	0.110	.
Sequence	577	APGGPYR|LG	0.049	.
Sequence	590	CPNATSK|IG	0.072	.
Sequence	603	MAWAVPK|DN	0.075	.
Sequence	607	VPKDNYK|NA	0.094	.
Sequence	638	GFHSDNK|TQ	0.143	.
Sequence	642	DNKTQMK|SL	0.088	.
Sequence	652	GDSNPQK|FT	0.074	.
Sequence	685	GLPQSGR|IV	0.142	.

Sequence	694	VDYMMQK|PG	0.080	.
Sequence	697	MMQKPGK|TG	0.074	.
Sequence	705	GTIVYQR|GV	0.035	.
Sequence	712	GVLLPQK|VW	0.113	.
Sequence	719	VWCASGR|SK	0.038	.
Sequence	721	CASGRSK|VI	0.081	.
Sequence	724	GRSKVIK|GL	0.069	.
Sequence	729	IKGLEQK|LI	0.149	.

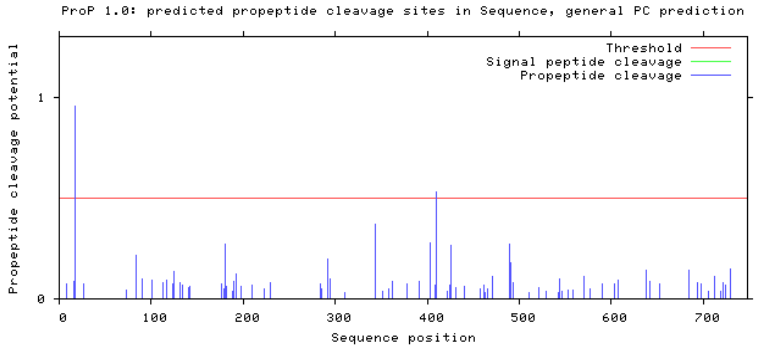


Table 23. The predicted truncated peptide derived from Human HA-RBD mutV3.3 after an hypothetical cleavage site cut. The molecular weight of the predicted peptide (36.172 KDa) is highly similar to that observed after IMAC purification and deglycosylation assays performed over the expressed Human HA RBD mutV3.3 (~35KDa) protein in *Pichia pastoris*. Parameter was calculated using the Bioinformatic tool ExpasyCompute pI/MWserver.

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

```
        10                  20                  30                  40                  50                  60
SQQAVIPNIG          SRPRIRNIPS          RISIYWTIVK          PGDILLINST          GNLIAPRGYF          KIRSGKSSIM

        70                  80                  90                 100                 110                 120
RSDAPIGGGG          SGGGGSGTRT          RGKLCPDCLN          CTDLDVALGR          PMCVGTTPSA          KASILHEVRP

       130                 140                 150                 160                 170                 180
VTSGCFPIMH          DRTKIRQLPN          LLRGYENIRL          STQNVIDAEK          APGGPYRLGT          SGSCPNATSK

       190                 200                 210                 220                 230                 240
IGFFATMAWA          VPKDNYKNAT          NPLTVEVPYI          CTEGEDQITV          WGFHSDNKTQ          MKSLYGDSNP

       250                 260                 270                 280                 290                 300
QKFTSSANGV          TTHYVSQIGD          FPDQTEDGGL          PQSGRIVVDY          MMQKPGKTGT          IVYQRGVLLP

       310                                                     320                                     330
QKVWCASGRS KVIKGLEQKL ISEEDLNHHH HHH
```

Theoretical pI/**:** 9.36 **Mw** /: **36172.18**

Table 24. The protease cleavage site as calculated for the Avian HA-RBD mutV3.7 obtained from the ProP 1.0 server. For comparison with the Human HA RBD mutV3.3, only the *Pichia pastoris* endogeneous Kex2 cleavage site (highlighted in green) is correctly predicted with a high probability value (0.975) within the Avian Ha RBD 3.7 peptide sequence. In the graph it is observed that there is a lack of additional internal cleavage sites above the threshold line for any of the calculated propeptide cleavage sites (PC, blue bars) beside the Kex2 signal peptide encoded by *Pichia pastoris* pPicZalphaA vector (In green).

```
ProP v.1.0b ProPeptide Cleavage Site Prediction
General PC cleavage site prediction (Arginine/Lysine residues)

  Avian HA RBD mutV3.7 Sequence
TIASIAAKEEGVSLEKRSVKGGFILKDCSVAGWLLGNPMCDEFLTAPEWSYIVEKDNPSNDLCYPGNFDNYEELKHLMSS      80
TNHMERIQIFPRNSWTNHNASAGVSSACPFNGGSSFFRNVVWLIKKNDVYRTLKRNYTNTNVEDLLIIWGVHHPNDATEQ     160
TKLYQNLNTYVSVGTSTLNQRSTPTIATRPQVNGQRGRMEFFWTILKSNDSIFFESTGNFIAPEYAYKIVKKGKSTVMKS     240
GGGGSGGGGSPLILKDCSVAGWLLGNPMCDEFLNVQEWSYIVEKDNPSNDLCYPGNFDNYEELKHLMSSTNHLERIQIFP     320
RNSWPNHDSSSGVSSACPFNGESSFFRNVVWLIKKNDVYRTLKRNYTNTNVEDLLIIWGVHHPNDATVQTKLYQNLNTYV     400
SVGTSTLNQRSIPTIATRPQVNGQRGRVEFFWTIL ESNDSIFFESTGNFIAPEYAYKVTKKGKSAVMKSGGGGSGGGGSP     480
TDLGQCGLLGTLIGPPQCDQFLKFDADLIIERREGTDVCYPGKFTNEESLRQILRRSGGIDKESMGFTYSGIRTNGTTSA     560
CRRSNPSFYAEMKWLLSNSDNATFPQMTKSYRNPRNKPALITWGVHHSGSATEQTKLYGSGDKLITVGSSKYLQSFTPSP     640
GARPQVNGQSGRIDFHWLLLDPNDTVTFTFNGAFIAPDRASFFRGESIGIQSKWCECRKKRRQRRRSLEQKLISEEDLNH     720
HHHHH                                                                               800
.............P.............................................................      80
...............................................................................    160
...............................................................................    240
...............................................................................    320
...............................................................................    400
...............................................................................    480
...............................................................................    560
...............................................................................    640
...............................................................................    720
.....                                                                              800

Propeptide cleavage sites predicted:  Arg(R)/Lys(K): 1
```

Name	Pos	Context v	Score	Pred
Sequence	8	IASIAAK|EE	0.071	.
Sequence	16	EGVSLEK|RS	0.081	.
Sequence	17	GVSLEKR|SV	0.975	*ProP*
Sequence	20	LEKRSVK|GG	0.082	.
Sequence	26	KGGFILK|DC	0.096	.
Sequence	55	WSYIVEK|DN	0.081	.
Sequence	75	DNYEELK|HL	0.061	.
Sequence	86	STNHMER|IQ	0.188	.
Sequence	92	RIQIFPR|NS	0.030	.
Sequence	118	GGSSFFR|NV	0.159	.
Sequence	125	NVVWLIK|KN	0.066	.
Sequence	126	VVWLIKK|ND	0.118	.
Sequence	131	KKNDVYR|TL	0.064	.
Sequence	134	DVYRTLK|RN	0.043	.
Sequence	135	VYRTLKR|NY	0.048	.
Sequence	162	DATEQTK|LY	0.068	.
Sequence	181	TSTLNQR|ST	0.311	.
Sequence	189	TPTIATR|PQ	0.034	.
Sequence	196	PQVNGQR|GR	0.082	.
Sequence	198	VNGQRGR|ME	0.066	.
Sequence	207	FFWTILK|SN	0.143	.
Sequence	228	APEYAYK|IV	0.064	.
Sequence	231	YAYKIVK|KG	0.068	.
Sequence	232	AYKIVKK|GK	0.073	.
Sequence	234	KIVKKGK|ST	0.063	.
Sequence	239	GKSTVMK|SG	0.062	.
Sequence	255	GSPLILK|DC	0.092	.
Sequence	284	WSYIVEK|DN	0.081	.
Sequence	304	DNYEELK|HL	0.061	.
Sequence	315	STNHLER|IQ	0.155	.
Sequence	321	RIQIFPR|NS	0.033	.
Sequence	347	GESSFFR|NV	0.124	.
Sequence	354	NVVWLIK|KN	0.066	.
Sequence	355	VVWLIKK|ND	0.118	.
Sequence	360	KKNDVYR|TL	0.064	.
Sequence	363	DVYRTLK|RN	0.043	.
Sequence	364	VYRTLKR|NY	0.048	.
Sequence	391	DATVQTK|LY	0.065	.
Sequence	410	TSTLNQR|SI	0.297	.
Sequence	418	IPTIATR|PQ	0.033	.
Sequence	425	PQVNGQR|GR	0.091	.
Sequence	427	VNGQRGR|VE	0.082	.
Sequence	457	APEYAYK|VT	0.079	.
Sequence	460	YAYKVTK|KG	0.060	.
Sequence	461	AYKVTKK|GK	0.082	.
Sequence	463	KVTKKGK|SA	0.064	.
Sequence	468	GKSAVMK|SG	0.057	.
Sequence	503	QCDQFLK|FD	0.061	.
Sequence	512	ADLIIER|RE	0.066	.
Sequence	513	DLIIERR|EG	0.047	.
Sequence	523	DVCYPGK|FT	0.117	.
Sequence	531	TNEESLR|QI	0.113	.
Sequence	535	SLRQILR|RS	0.034	.
Sequence	536	LRQILRR|SG	0.153	.
Sequence	542	RSGGIDK|ES	0.073	.
Sequence	553	FTYSGIR|TN	0.050	.
Sequence	562	GTTSACR|RS	0.042	.
Sequence	563	TTSACRR|SN	0.198	.
Sequence	573	SFYAEMK|WL	0.059	.
Sequence	589	TFPQMTK|SY	0.080	.
Sequence	592	QMTKSYR|NP	0.044	.
Sequence	595	KSYRNPR|NK	0.078	.
Sequence	597	YRNPRNK|PA	0.069	.
Sequence	616	SATEQTK|LY	0.089	.
Sequence	623	LYGSGDK|LI	0.086	.
Sequence	631	ITVGSSK|YL	0.105	.
Sequence	643	TPSPGAR|PQ	0.054	.
Sequence	652	VNGQSGR|ID	0.044	.
Sequence	679	AFIAPDR|AS	0.033	.
Sequence	684	DRASFFR|GE	0.055	.

Sequence	693	SIGIQSK|WC	0.082	.
Sequence	698	SKWCECR|KK	0.035	.
Sequence	699	KWCECRK|KR	0.073	.
Sequence	700	WCECRKK|RR	0.060	.
Sequence	701	CECRKKR|RQ	0.101	.
Sequence	702	ECRKKRR|QR	0.056	.
Sequence	704	RKKRRQR|RR	0.053	.
Sequence	705	KKRRQRR|RS	0.105	.
Sequence	706	KRRQRRR|SL	0.086	.
Sequence	711	RRSLEQK|LI	0.107	.

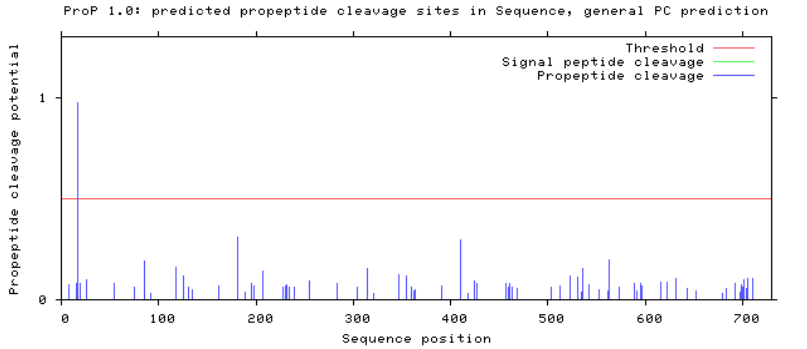


Table 25. Amino acids mutagenesis on the predicted pre-sequence and cleavage sites (context) and its possible outcome. The betterbest predicted substitutions are those located within the pre-sequence (in bold, X4X3X2X1, where X is any amino acid residue which is necessary within the cleavage context, and the cleavage site (B1B2, where B is a basic residue) of the Kex2 region. The underlined residue(s) in the predicted sequences represent the mutated site.

Mutagenesis substitution predictor	Position	Context (**X4X3X2X1**/**B1B2**)	Score	Prediction
Wt	25	ITVSTKR|SQ	0.527	*ProP*
Mut1 (I/A)	25	ATVSTKR|SQ	0.461	.
Mut2 (T/A)	25	IAVSTKR|SQ	0.758	*ProP*
Mut3 (V/A)	25	ITASTKR|SQ	0.682	*ProP*
Mut4 (S/A)	25	ITVATKR|SQ	0.256	.
Mut5 (T/A)	25	ITVSAKR|SQ	0.417	.
Mut6 (K/A)	**25**	I**TVSTAR|**SQ	**0.063**	.
Mut7 (R/A)	**24**	RI**TVSTK|A**S	**0.059**	.
Mut8 (S/A)	25	ITVSTKR|AQ	0.670	*ProP*
Mut9 (Q/A)	25	ITVSTKR|SA	0.463	.
Mut10 (K/R)	**25**	I**TVSTRR|**SQ	**0.140**	.
Mut11 (R/K)	25	ITVSTKK|SQ	0.312	.
Mut12 (K/R)/(R/K)	**25**	I**TVSTRK|**SQ	**0.100**	.

4.0. DISCUSSION

The results of the first attempt to induce the expression of HA-RBD multimers lacking additional tagging sequences in *Pichia pastoris* expression system, indicated that our protein of interest, specifically, the Avian HA RBD V3.4 was detected in the samples analyzed by MASS-Spectroscopy and isolated by SDS-PAGE, anion-exchange chromatography, and size exclusion chromatography. Nevertheless, Mass-Spectroscopy results for this molecule identified additional proteins, of Yeast origin, as potential contaminants leading to miss-identification of the selected band (chromatography **Figure 20**). This suggests the co-isolation of more than one *Pichia pastoris* protein during purification of the recombinant protein (**Appendix Table A1**). The mass spectroscopy also indicates possible proteins of human origin (indicating cross contamination during sample preparation) and proteins from the media source in addition to our expected HA-RBD V3.4 peptides, indicating co-purification of peptides isolated by the chromatography methods. Additionally, the N-terminal sequencing chromatograms of the derivatized residues did not show a single, clear signaling peak as expected for an absolute sequence determination suggesting the presence of multiples sequences (82). Since subsequent analysis has identified heterogeneous and highly glycosylated His-Tag versions of the expressed protein, it is possible that the first attempt at purification isolated heterogeneous forms from Avian HA RBD V3.4 glycosylated antigen that may be co-purified in the analysed sample (**Figure 20, Figure 21**). Ferret anti-H5N1 polyclonal immunodetection of the 200 mM NaCl anion exchange chromoatography eluent (**Figure 24**) of the Avian HA-RBD V3.4 samples identified a high molecular size region. This indicated that the expressed Avian HA-RBD V3.4

could be present at a higher molecular weight than was expected (72-78 KDa). This was corroborated later in the expression strategy using the His-Tag versions (**Figure 32**).

At the level of quantified protein expression, the extracellular soluble protein recovered following IMAC-dependent purification of the Human HA-RBD mutV3.3 transformed yeast (105 □g/mL) had a higher yield than Avian HA-RBD mutV3.4 (63 □g/ml) and the Avian HA-RBD mutV3.6 (22 □g/mL) yeast (**Figure 30**). Extracelullar expression analysis of the Avian HA-RBD mutV3.7 proteins remains to be done.

The Human HA-RBD mutV3.3 is expressed primarily as an heterogeneous hyper-glycosylated and proteolyzed form (~36KDa)(**Figure 32**) after 48 h of methanol induction. This makes this contruct currently unsuitable to initiate a study to measure responses to Human trivalent RBD immunization. However, there is an enzymatically de-glycosylated form of the protein band, in the range of ~87 KDa, that could represent a partially de-glycosylated form, or alternatively this same protein band may be a full-length HA RBD mutV3.3 polypeptide. Staining for carbohydrate side chains would help distinguish these possibilities. If the latter assumption is correct, it could be possible to adjust the conditions used for protein induction to increase the protein yield of the non-cleaved HA-RBD mutV3.3 peptide For example, the alkalinity of the protein expression media could be increased to pH 8.0, the time of expression couild be decreased from 48h to 24 h, or a Protease deficient *Pinchia pastoris* strain (Pink Strain) could be used to express the transformed gene, as possible solutions. On the other hand, the Avian HA-RBD mutV3.4 and Avian HA RBD mutV3.6 transformed yeast expressed full length, non-truncated forms as assessed by the de-glycosylation assays. However, these

recombinant proteins were expressed at lower yields compared to the Human version mutV3.3 as detected in immunoblot analysis using the anti-his monoclonal antibody. Optimization of the conditions to increase expression of the recombinant protein, like re-transformation of the yeast with the expression DNA vector to increase copy number, fermentation scale-up, optimization of the media composition, increasing the final density in the media at the time of induction (O.D=4 instead of O.D=1), increasing the pH of the media, decreasing the final temperature of induction to 25^{o}C from 30^{o}C, increasing the final methanol percentage in the media (from 0.5 % to 2% vol/vol), to mention a few possibilities, could improve the protein yield.

In terms of protein characterization, sequencing results of both the Avian and Human versions did not show base modifications during the cloning process. The expression products of these transformed plasmids were recognized by a specific anti-6-His Tag monoclonal antibody suggesting that the group of peptides are in frame with the 6Histidine tag present in the *Pichia pastoris* vector used to express the Avian HA-RBD mutV3.4 and the Avian HA-RBD mutV3.6. The Human HA-RBD mutV3.3 and Avian HA-RBD mutV3.7 constructs remains to be sequenced, since additional primers are required for complete gene sequencing. ,

Additional characterization using ferret specific polyclonal anti-influenza virus antibodies that were created to recognize H7 and H5 subtype epitopes are also able to detect immunoreactive proteins in the non-de-glycosylated Avian HA RBD mutV3.4 and Avian HA-RBD mutV3.6 samples. This suggests there is an interaction between the polyclonal antibodies and the linear Avian HA-RBD epitopes (the samples were reduced

in DTT before being loaded on the SDS-PAGE gel to further disrupt tertiary structure). Based on the dissimilarities in molecular size observed between the different proteins that are identified, depending on the type of anti-sera used (ie H5 or H7 subtype directed antibodies), it is likely that the high molecular size 265 KDa protein that is detected with with the FR-231 antiserum (H5 subtyping) may represent a highly glycosylated molecule that includes the RBD1 and RBD2 regions (H5 subtype). However, this protein was not recognized by the FR-1088 antiserum (H7 subtype), although a smaller 165 kDa protein was detect with this antibody and could represent a glycosylation-free form of the RBD3 (H7 subtype) region in the multimeric HA-RBD protein.

In the case of the Human HA-RBD mutV3.3, only the specific ferret polyclonal antibodies raised against Influenza B virus (Yamagata Lineage) were able to recognize the purified protein. The ferret polyclonal antibodies raised against the influenza A/H1N1/07/California 2009 pandemic virus did not recognize the protein. This indicates that the protein had been proteolytically cleaved to generate the N-terminal RBD1 in frame with the His6 purification tag that did not contain the remaining RBD domains. , derived from the H1 subtype and part of the RBD2 derived from the H3 subtype of the mutV3.3 version.

Further experiments are required to confirm that the mass spectroscopy and N-terminal sequencing of the Avian and Human HA-RBD construct fully demonstrate their identity. These experiments will give valuable information that is required in order to proceed with animal immunization studies. N-Glycosylation, assessed by Endo H glycosidase, appears to be the main type of carbohydrate added by the *Pichia pastoris*

expression system for all ofsd the HA-RBD versions evaluated (mutV3.3, mutV3.4, mutV3.6). The characteristic N-linked consensus sequence was used to calculate the potential N-glycosylated forms that might be occurring in the expressed proteins (mutV3.3, mutV3.4 and applies for mutV3.6, as well). This analysis might also demonstrate which residues to modify in order to eliminate endogenous glycosylation of the recombinant proteins (**Figure 35, Figure 36**).

The experiments that examined intracellular expression of the HA-RBDs by the *Pinchia pastoris* transformants did not generate a confirmed candidate. The yeast were transformed with HA-RBD DNA constructs that contained a HisT tag sequence but preliminary results did not show any of the recovered proteins were positive in immunoblots with Anti-6-Histidine antibodies. Nevertheless, to discard the possibility of using intracellular expression and protein recovery conditions is premature. These conditions should be modified since these protein products may be useful since they represent low-glycosylated forms not targeted to the secretory pathway and therefore have the potential to induce a broader immune responses as observed in de-glycosylated hemagglutinin molecules evaluated by other influenza research groups (61,63,65,83–91).

<u>His-Tagged Human HA-RBD mutV3.3 and its post-translational process. A tale for Avian versions, as well.</u>

The truncation of the human HA-RBD mutV3.3, likely by proteolysis, might also be minimized in an intracellular expression system. Most of the Kex2 activity, which likely mediates this cleavage, is found in the endoplasmic reticulum where the

recombinant protein is exposed during translocation of the exported protein (92–94). Intracellular expressed proteins would not encounter the Kex2 protease.

In *Pichia pastoris,* Kex2 is a highly abundant subtilisin/kexin-like proprotein convertase (PC) member of the proteolytic enzyme family that is responsible for post-translational processing and translocation through the secretory pathway. Kex2 processes the signal peptide of the transported proteins and is localized to the transmembrane Golgi network (92) and endocytic compartment (95). Based on the expression pattern and the truncation of the Human HA-RBD mutV3.3 and the results of the protein analysis that detected an internal PC signal, the results are consistent with the identification of the de-glycosylated ~37 KDa Human HA-RBD mutV3.3 truncated peptide (**Figure 32**) as resulting from the presence of an internal sequence of the Influenza virus H3 subtype HA-RBD subunit 2 (**Table 20, Table 21**) Furthermore, the predicted data for the Avian HA-RBD versions (**Table 22**) are consistent with the lack of observed major proteolysis degradation as determined by SDS-PAGE for the IMAC-purified and de-glycosylated protein (**Figure 32**). Results compared the observed and predicted data (**Table 21**) indicate that is possible to correct the truncated peptide within the Human HA-RBD ITVSTKR|SQ (0.527) cleavage sequence to generate a mutated sequence that avoids cleavage by the Kex2 protease. The predicted substitution analysis suggest that in order to avoid post-translation processing of the internal predicted Kex2 cleavage site, the human H3 HA RBD subunit may require the use of alanine-scanning mutagenesis (96,97). In particular, mutation of the Mut7 (R/A)(**0.059**) and Mut6 (K/A)(**0.063**) sequences might be best first options to evaluate in developing new experimental Human HA RBD mutV3.3 mutants. Two additional mutants with

basic/basic amino-acids substitution sequences, the double basic mutant Mut12 (K/R)/(R/K)(**0.100**) and the Mut10 (K/R)(**0.140**), represent the next alternative mutations to evaluate the effect on truncation repair and full-size expression of Human HA RBD sequences. However, even though the positive/positive charge substitution required in these mutations may seem not to be of big concern, some previous structure-modelled analysis has shown these changes could be pertinent since the amino acid charges in the antigenic sites have been correlated with changes in the antigenic cluster in influenza virus (98). Whether thees substitutions could impact the antigenic recognition sites for the well characterized Influenza H3 region (98–100), needs to be evaluated experimentally. Undoubtedly, the experimental analysis of these new Human HA RBD mutants could help us to shape a better design for Human HA RBDs molecules. This ultimately could impact their evaluation as immunogenic molecules at least, when the use of Human H3 subtype RBD's are used as potential vaccine candidates in a *Pichia pastoris* expression platform.

There is some evidence from studies using mutational analysis, removal of carbohydrates, or carbohydrate shielding that glycosylation of vaccinating influenza hemagglulin proteins may enhance the production of functional antibodies (83–86,88,89,101). However, further experimental evidence is required to test the potential effect of N-glycosylation on vaccine design, proper immune system recognition, and production of neutralizing antibodies. Site-directed mutagenesis substitution at residues within the glycosylation motif N-X-S/T, with the preference to substitute the asparagine residue position with aspartate or to substitute lysine at the serine residue position of the motif (83) can be performed. These studies should reveal the effect of substituting

for any of the seven structural permitted N-glycosylation motifs in the Avian HA-RBD mutV3.4 and mutV3.6 versions or any of the nine consensus N-glycosylation domains in the full length Human HA-RBD mutV3.3 or any of the five consensus domains contained in the truncated Human HA-RBD mutV3.3 (**Figure 35, Figure 36E**). Similar studies have shown that changes in glycosylation can effect the Receptor Binding Domain derived from MERS-CoV (65,102).

Even more, this study demonstrates that *Pichia pastoris* could provide an excellent expression system for work related to the expression of HA-RBD designs where the presence rather than absence of glycosylation may be important for biological function For example, HA-RBD can be designed an selected by directed-evolution in order to have the highest binding affinities for sialic acid moieties. The generated molecules could then be serving as competitive inhibitors for the binding of wild type hemagglutinins required for virus recognition, attachment and infection in the respiratory epithelium. Therefore, hyper-glycosylated and functionally-active forms of HA-RBD expressed in *Pichia pastoris* could serve as a shielding strategy against the activation of the immune response without compromising the biological intended function.

FUTURE DIRECTIONS:

1. Site-directed mutagenesis for removal of N-Glycosylation sites are necessary to determine if these point mutation can generate homogeneous polypeptides rather than t the multiple variations that carbohydrate heterogeneity may be present after protein isolation and immunoblot analysis of the HA-RBD multimeric genes.

2. Site-directed mutagenesis of the Human HA-RBD version (H3 subtype) to alter the putative internal peptidase cleavage site should facilitate recovery of the full-size recombinant protein and will demonstrate the suitability of human HA-RBD version for use as a vaccine. This version was shown to have a higher yield which is the main purpose for using *Pichia pastoris* as a heterologous expression system.

3. Optimization of the process for expressing and purifying the intracellular HA-RBD versions could overcome issues with protein cleavage and glyosylation that result in the heterogeneous expression of the hyperglycosylated HA-RBD version evaluated to date.

4. Tests of functional activity and immunogenicity of the N-glycosylated forms, deglycosylated forms, and mut Deglycosylated forms will be required in order to confirm that the HA-RBD are properly folded, that they can interact with cellular

surface receptors, and that they can induce appropriate neutralizing antibody production in vaccinated animal experiments. The utility of the Human and Avian HA-RBD versions could be applied to different platforms of generating immunity (for example, DNA vaccines)that could be demonstrated if all the aforementioned experiments were shown to be an effective and could indicate a strategy to generate novel Influenza vaccines.

Bibliography

1. Medina RA, García-Sastre A. Influenza A viruses: new research developments. Nat Rev Microbiol. 2011 Aug;9(8):590–603.

2. Statement on Seasonal Influenza Vaccine for 2014-2015 - Public Health Agency of Canada [Internet]. Available from: http://www.phac-aspc.gc.ca/naci-ccni/flu-grippe-eng.php

3. Weekly influenza reports [Internet]. Available from: http://healthycanadians.gc.ca/diseases-conditions-maladies-affections/disease-maladie/flu-grippe/surveillance/fluwatch-reports-rapports-surveillance-influenza-eng.php

4. Review of the 2014-2015 influenza season in the northern hemisphere. Wkly Epidemiol Rec. 2015 Jun 5;90(23):281–296.

5. Kapoor S, Dhama K. Insight into Influenza Viruses of Animals and Humans. Cham: Springer International Publishing; 2014.

6. Taubenberger JK, Kash JC. Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe. 2010 Jun 25;7(6):440–451.

7. Sriwilaijaroen N, Suzuki Y. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. Proc Jpn Acad, Ser B, Phys Biol Sci. 2012;88(6):226–249.

8. Bourmakina SV, García-Sastre A. Reverse genetics studies on the filamentous morphology of influenza A virus. J Gen Virol. 2003 Mar;84(Pt 3):517–527.

9. Itoh Y, Shinya K, Kiso M, Watanabe T, Sakoda Y, Hatta M, et al. In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. Nature. 2009 Aug 20;460(7258):1021–1025.

10. Goldsmith CS, Metcalfe MG, Rollin DC, Shieh WJ, Paddock CD, Xu X, et al. Ultrastructural characterization of pandemic (H1N1) 2009 virus. Emerging Infect Dis. 2011 Nov;17(11):2056–2059.

11. Campbell PJ, Danzy S, Kyriakis CS, Deymier MJ, Lowen AC, Steel J. The M segment of the 2009 pandemic influenza virus confers increased neuraminidase activity, filamentous morphology, and efficient contact transmissibility to A/Puerto Rico/8/1934-based reassortant viruses. J Virol. 2014 Apr;88(7):3802–3814.

12. Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature. 1981 Jan 29;289(5796):366–373.

13. Wilson IA, Cox NJ. Structural basis of immune recognition of influenza virus hemagglutinin. Annu Rev Immunol. 1990;8:737–771.

14. Matsubara T, Onishi A, Saito T, Shimada A, Inoue H, Taki T, et al. Sialic acid-mimic peptides as hemagglutinin inhibitors for anti-influenza therapy. J Med Chem. 2010 Jun 10;53(11):4441–4449.

15. Krammer F, Palese P. Advances in the development of influenza virus vaccines. Nat Rev Drug Discov. 2015 Mar;14(3):167–182.

16. Cho KJ, Lee JH, Hong KW, Kim SH, Park Y, Lee JY, et al. Insight into structural diversity of influenza virus haemagglutinin. J Gen Virol. 2013 Aug;94(Pt 8):1712–1722.

17. Connor RJ, Kawaoka Y, Webster RG, Paulson JC. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. Virology. 1994 Nov 15;205(1):17–23.

18. Sieczkarski SB, Whittaker GR. Influenza virus can enter and infect cells in the absence of clathrin-mediated endocytosis. J Virol. 2002 Oct;76(20):10455–10464.

19. Perez JT, Varble A, Sachidanandam R, Zlatev I, Manoharan M, García-Sastre A, et al. Influenza A virus-generated small RNAs regulate the switch from transcription to replication. Proc Natl Acad Sci U S A. 2010 Jun 22;107(25):11525–11530.

20. Chenavas S, Estrozi LF, Slama-Schwok A, Delmas B, Di Primo C, Baudin F, et al. Monomeric nucleoprotein of influenza A virus. PLoS Pathog. 2013 Mar 28;9(3):e1003275.

21. Roberts PC, Lamb RA, Compans RW. The M1 and M2 proteins of influenza A virus are important determinants in filamentous particle formation. Virology. 1998 Jan 5;240(1):127–137.

22. Varghese JN, Colman PM. Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 A resolution. J Mol Biol. 1991 Sep 20;221(2):473–486.

23. França M, Stallknecht DE, Howerth EW. Expression and distribution of sialic acid influenza virus receptors in wild birds. Avian Pathol. 2013 Feb 1;42(1):60–71.

24. König R, Stertz S, Zhou Y, Inoue A, Hoffmann HH, Bhattacharyya S, et al. Human host factors required for influenza virus replication. Nature. 2010 Feb 11;463(7282):813–817.

25. Matrosovich MN, Gambaryan AS, Teneberg S, Piskarev VE, Yamnikova SS, Lvov DK, et al. Avian influenza A viruses differ from human viruses by recognition

of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. Virology. 1997 Jun 23;233(1):224–234.

26. Matrosovich M, Tuzikov A, Bovin N, Gambaryan A, Klimov A, Castrucci MR, et al. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J Virol. 2000 Sep;74(18):8502–8512.

27. Schrauwen EJ, Fouchier RA. Host adaptation and transmission of influenza A viruses in mammals. Emerging microbes & infections. 2014 Feb 12;3(2):e9.

28. Zeng H, Goldsmith CS, Maines TR, Belser JA, Gustin KM, Pekosz A, et al. Tropism and infectivity of influenza virus, including highly pathogenic avian H5N1 virus, in ferret tracheal differentiated primary epithelial cell cultures. J Virol. 2013 Mar;87(5):2597–2607.

29. Trifonov V, Khiabanian H, Rabadan R. Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. N Engl J Med. 2009 Jul 9;361(2):115–119.

30. Writing Committee of the WHO Consultation on Clinical Aspects of Pandemic (H1N1) 2009 Influenza, Bautista E, Chotpitayasunondh T, Gao Z, Harper SA, Shaw M, et al. Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection. N Engl J Med. 2010 May 6;362(18):1708–1719.

31. Wong SS, Webby RJ. Traditional and new influenza vaccines. Clin Microbiol Rev. 2013 Jul;26(3):476–492.

32. Seasonal Influenza Vaccine & Total Doses Distributed | Health Professionals | Seasonal Influenza (Flu) [Internet]. Available from: http://www.cdc.gov/flu/professionals/vaccination/vaccinesupply.htm

33. Russell K, Blanton L, Kniss K, Mustaquim D, Smith S, Cohen J, et al. Update: influenza activity--United states, october 4, 2015-February 6, 2016. MMWR Morb Mortal Wkly Rep. 2016 Feb 19;65(6):146–153.

34. Franco-Paredes C, Carrasco P, Preciado J. The first influenza pandemic in the new millennium: lessons learned hitherto for current control efforts and overall pandemic preparedness. J Immune Based Ther Vaccines. 2009;7(1):2.

35. Petsch B, Schnee M, Vogel AB, Lange E, Hoffmann B, Voss D, et al. Protective efficacy of in vitro synthesized, specific mRNA vaccines against influenza A virus infection. Nat Biotechnol. 2012 Dec;30(12):1210–1216.

36. Chua JV, Chen WH. Bench-to-bedside review: vaccine protection strategies during pandemic flu outbreaks. Crit Care. 2010 Apr 16;14(2):218.

37. Taxonomy browser (Influenza A virus (A/reassortant/NYMC X-181(A/NYMC X-157 x California/07/2009)(H1N1))) [Internet]. Available from: http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=1184516

38. Pearson H. Flu vaccine shortage looms. news@nature. 2004 Oct 4;

39. Bodles-Brakhop AM, Heller R, Draghia-Akli R. Electroporation for the delivery of DNA-based vaccines and immunotherapeutics: current clinical developments. Mol Ther. 2009 Apr;17(4):585–592.

40. Zheng L, Wang F, Yang Z, Chen J, Chang H, Chen Z. A single immunization with HA DNA vaccine by electroporation induces early protection against H5N1 avian influenza virus challenge in mice. BMC Infect Dis. 2009 Feb 12;9:17.

41. Yan J, Villarreal DO, Racine T, Chu JS, Walters JN, Morrow MP, et al. Protective immunity to H7N9 influenza viruses elicited by synthetic DNA vaccine. Vaccine. 2014 May 19;32(24):2833–2842.

42. Scott VL, Patel A, Villarreal DO, Hensley SE, Ragwan E, Yan J, et al. Novel synthetic plasmid and Doggybone DNA vaccines induce neutralizing antibodies and provide protection from lethal influenza challenge in mice. Hum Vaccines Immunother. 2015;11(8):1972–1982.

43. Muthumani K, Block P, Flingai S, Muruganantham N, Chaaithanya IK, Tingey C, et al. Rapid and Long-Term Immunity Elicited by DNA-Encoded Antibody Prophylaxis and DNA Vaccination Against Chikungunya Virus. J Infect Dis. 2016 Aug 1;214(3):369–378.

44. Chung C, Ugen KE, Sardesai NY, Weiner DB, Muthumani K. Protocols for Developing Novel Chikungunya Virus DNA Vaccines. Methods Mol Biol. 2016;1426:311–332.

45. Azizi A, Diaz-Mitoma F. Viral peptide immunogens: current challenges and opportunities. J Pept Sci. 2007 Dec;13(12):776–786.

46. Gottlieb T, Ben-Yedidia T. Epitope-based approaches to a universal influenza vaccine. J Autoimmun. 2014 Nov;54:15–20.

47. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov. 2007 May;6(5):404–414.

48. Samayoa L, Diaz-Mitoma F, Azizi A. Characterization of a branched lipopeptide candidate vaccine against influenza A/Puerto Rico 8/34 which is recognized by human B and T-cell immune responses. Virol J. 2011 Jun 16;8:309.

49. Krammer F, Palese P. Universal influenza virus vaccines: need for clinical trials. Nat Immunol. 2014 Jan;15(1):3–5.

50. Chen W, Feng Y, Wang Y, Zhu Z, Dimitrov DS. Fusion proteins of HIV-1 envelope glycoprotein gp120 with CD4-induced antibodies showed enhanced binding to CD4 and CD4 binding site antibodies. Biochem Biophys Res Commun. 2012 Sep 7;425(4):931–937.

51. Cregg JM, Cereghino JL, Shi J, Higgins DR. Recombinant protein expression in Pichia pastoris. Mol Biotechnol. 2000 Sep;16(1):23–52.

52. Tschopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR. Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. Nucleic Acids Res. 1987 May 11;15(9):3859–3876.

53. Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J. GAP promoter library for fine-tuning of gene expression in Pichia pastoris. Appl Environ Microbiol. 2011 Jun;77(11):3600–3608.

54. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene. 1997 Feb 20;186(1):37–44.

55. Frenzel A, Hust M, Schirrmann T. Expression of recombinant antibodies. Front Immunol. 2013 Jul 29;4:217.

56. Athmaram TN, Saraswat S, Santhosh SR, Singh AK, Suryanarayana WS, Priya R, et al. Yeast expressed recombinant Hemagglutinin protein of novel H1N1 elicits neutralising antibodies in rabbits and mice. Virol J. 2011 Nov 29;8:524.

57. Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol. 2014 Jun;98(12):5301–5317.

58. Doerner A, Rhiel L, Zielonka S, Kolmar H. Therapeutic antibody engineering by high efficiency cell screening. FEBS Lett. 2014 Jan 21;588(2):278–287.

59. Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, et al. Production of complex human glycoproteins in yeast. Science. 2003 Aug 29;301(5637):1244–1246.

60. Grinna LS, Tschopp JF. Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, Pichia pastoris. Yeast. 1989 Apr;5(2):107–115.

61. Yang YL, Chang SH, Gong X, Wu J, Liu B. Expression, purification and characterization of low-glycosylation influenza neuraminidase in α-1,6-mannosyltransferase defective Pichia pastoris. Mol Biol Rep. 2012 Feb;39(2):857–864.

62. Wildt S, Gerngross TU. The humanization of N-glycosylation pathways in yeast. Nat Rev Microbiol. 2005 Feb;3(2):119–128.

63. Lin Q, Yang K, He F, Jiang J, Li T, Chen Z, et al. Production of Influenza Virus HA1 Harboring Native-Like Epitopes by Pichia pastoris. Appl Biochem Biotechnol. 2016 Apr 4;

64. DuBois RM, Aguilar-Yañez JM, Mendoza-Ochoa GI, Oropeza-Almazán Y, Schultz-Cherry S, Alvarez MM, et al. The receptor-binding domain of influenza virus hemagglutinin produced in Escherichia coli folds into its native, immunogenic structure. J Virol. 2011 Jan;85(2):865–872.

65. Wang CC, Chen JR, Tseng YC, Hsu CH, Hung YF, Chen SW, et al. Glycans on influenza hemagglutinin affect receptor binding and immune response. Proc Natl Acad Sci U S A. 2009 Oct 27;106(43):18137–18142.

66. Phage Display: A Laboratory Manual [Internet]. Available from: http://www.cshlpress.com/default.tpl?action=full&--eqskudatarq=520

67. Xu W, Zheng M, Zhou F, Chen Z. Long-term immunogenicity of an inactivated split-virion 2009 pandemic influenza A H1N1 virus vaccine with or without aluminum adjuvant in mice. Clin Vaccine Immunol. 2015 Mar;22(3):327–335.

68. Aguilar-Yáñez JM, Portillo-Lara R, Mendoza-Ochoa GI, García-Echauri SA, López-Pacheco F, Bulnes-Abundis D, et al. An influenza A/H1N1/2009 hemagglutinin vaccine produced in Escherichia coli. PLoS ONE. 2010 Jul 22;5(7):e11694.

69. Impagliazzo A, Milder F, Kuipers H, Wagner MV, Zhu X, Hoffman RM, et al. A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. Science. 2015 Sep 18;349(6254):1301–1306.

70. Lee PS, Zhu X, Yu W, Wilson IA. Design and Structure of an Engineered Disulfide-Stabilized Influenza Virus Hemagglutinin Trimer. J Virol. 2015 Jul;89(14):7417–7420.

71. 2013-2014 Influenza Season | Seasonal Influenza (Flu) | CDC [Internet]. Available from: http://www.cdc.gov/flu/pastseasons/1314season.htm

72. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun;10(6):845–858.

73. Wu S, Letchworth GJ. High efficiency transformation by electroporation of Pichia pastoris pretreated with lithium acetate and dithiothreitol. BioTechniques. 2004 Jan;36(1):152–154.

74. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc. 2006;1(6):2856–2860.

75. Hahne H, Pachl F, Ruprecht B, Maier SK, Klaeger S, Helm D, et al. DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. Nat Methods. 2013 Oct;10(10):989–991.

76. Killian ML. Hemagglutination assay for the avian influenza virus. Methods Mol Biol. 2008;436:47–52.

77. Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. Protein Eng Des Sel. 2004 Jan;17(1):107–112.

78. Dreyfus C, Laursen NS, Kwaks T, Zuijdgeest D, Khayat R, Ekiert DC, et al. Highly conserved protective epitopes on influenza B viruses. Science. 2012 Sep 14;337(6100):1343–1348.

79. Zhu X, Viswanathan K, Raman R, Yu W, Sasisekharan R, Wilson IA. Structural Basis for a Switch in Receptor Binding Specificity of Two H5N1 Hemagglutinin Mutants. Cell Rep. 2015 Nov 24;13(8):1683–1691.

80. Russell RJ, Gamblin SJ, Haire LF, Stevens DJ, Xiao B, Ha Y, et al. H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. Virology. 2004 Aug 1;325(2):287–296.

81. Wu Y, Cho M, Shore D, Song M, Choi J, Jiang T, et al. RCSB PDB - 4R8W: Crystal structure of H7 hemagglutinin from A/Anhui/1/2013 in complex with a neutralizing antibody CT149 Structure Summary Page. Nat Commun [Internet]. 2015; Available from: http://www.rcsb.org/pdb/explore/explore.do?structureId=4R8W

82. Reim DF, Speicher DW. N-terminal sequence analysis of proteins and peptides. Curr Protoc Protein Sci. 2001 May;Chapter 11:Unit 11.10.

83. Medina RA, Stertz S, Manicassamy B, Zimmermann P, Sun X, Albrecht RA, et al. Glycosylations in the globular head of the hemagglutinin protein modulate the virulence and antigenic properties of the H1N1 influenza viruses. Sci Transl Med. 2013 May 29;5(187):187ra70.

84. Tate MD, Job ER, Deng YM, Gunalan V, Maurer-Stroh S, Reading PC. Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. Viruses. 2014 Mar 14;6(3):1294–1316.

85. Sun X, Jayaraman A, Maniprasad P, Raman R, Houser KV, Pappas C, et al. N-linked glycosylation of the hemagglutinin protein influences virulence and

antigenicity of the 1918 pandemic and seasonal H1N1 influenza A viruses. J Virol. 2013 Aug;87(15):8756–8766.

86. Job ER, Deng YM, Barfod KK, Tate MD, Caldwell N, Reddiex S, et al. Addition of glycosylation to influenza A virus hemagglutinin modulates antibody-mediated recognition of H1N1 2009 pandemic viruses. J Immunol. 2013 Mar 1;190(5):2169–2177.

87. Das SR, Puigbò P, Hensley SE, Hurt DE, Bennink JR, Yewdell JW. Glycosylation focuses sequence variation in the influenza A virus H1 hemagglutinin globular domain. PLoS Pathog. 2010 Nov 24;6(11):e1001211.

88. Skehel JJ, Stevens DJ, Daniels RS, Douglas AR, Knossow M, Wilson IA, et al. A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. Proc Natl Acad Sci U S A. 1984 Mar;81(6):1779–1783.

89. Zhang Y, Zhu J, Li Y, Bradley KC, Cao J, Chen H, et al. Glycosylation on hemagglutinin affects the virulence and pathogenicity of pandemic H1N1/2009 influenza A virus in mice. PLoS ONE. 2013 Apr 24;8(4):e61397.

90. Tsuchiya E, Sugawara K, Hongo S, Matsuzaki Y, Muraki Y, Li ZN, et al. Effect of addition of new oligosaccharide chains to the globular head of influenza A/H2N2 virus haemagglutinin on the intracellular transport and biological activities of the molecule. J Gen Virol. 2002 May;83(Pt 5):1137–1146.

91. Wiley DC, Wilson IA, Skehel JJ. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature. 1981 Jan 29;289(5796):373–378.

92. Redding K, Holcomb C, Fuller RS. Immunolocalization of Kex2 protease identifies a putative late Golgi compartment in the yeast Saccharomyces cerevisiae. J Cell Biol. 1991 May;113(3):527–538.

93. Bader O, Krauke Y, Hube B. Processing of predicted substrates of fungal Kex2 proteinases from Candida albicans, C. glabrata, Saccharomyces cerevisiae and Pichia pastoris. BMC Microbiol. 2008 Jul 14;8:116.

94. Daly R, Hearn MT. Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. J Mol Recognit. 2005 Apr;18(2):119–138.

95. Blanchette JM, Abazeed ME, Fuller RS. Cell-free reconstitution of transport from the trans-golgi network to the late endosome/prevacuolar compartment. J Biol Chem. 2004 Nov 19;279(47):48767–48773.

96. Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science. 1989 Jun 2;244(4908):1081–1085.

97. Lefèvre F, Rémy MH, Masson JM. Alanine-stretch scanning mutagenesis: a simple and efficient method to probe protein structure and function. Nucleic Acids Res. 1997 Jan 15;25(2):447–448.

98. Sun H, Yang J, Zhang T, Long LP, Jia K, Yang G, et al. Using sequence data to infer the antigenicity of influenza virus. MBio. 2013 Jul 2;4(4).

99. Gerhard W, Yewdell J, Frankel ME, Webster R. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. Nature. 1981 Apr 23;290(5808):713–717.

100. Stray SJ, Pittman LB. Subtype- and antigenic site-specific differences in biophysical influences on evolution of influenza virus hemagglutinin. Virol J. 2012 May 8;9:91.

101. Khurana S, Verma S, Verma N, Crevar CJ, Carter DM, Manischewitz J, et al. Bacterial HA1 vaccine against pandemic H5N1 influenza virus: evidence of oligomerization, hemagglutination, and cross-protective immunity in ferrets. J Virol. 2011 Feb;85(3):1246–1256.

102. Chen WH, Du L, Chag SM, Ma C, Tricoche N, Tao X, et al. Yeast-expressed recombinant protein of the receptor-binding domain in SARS-CoV spike protein with deglycosylated forms as a SARS vaccine candidate. Hum Vaccines Immunother. 2014;10(3):648–658.

Appendix Table 1. Full- set of peptides predicted for the MASS-Spec Results.

DB	Accession	Score	Mass	Matches	Pep(sig)	Sequences	Seq(sig)	emPAI	Description
Client_Sequences	1::CUSTOM_HA3.4	1080	75823	109	53	28	18	2.46	CUSTOM jaguilaryanez_amric.ca HA3.4
SwissProt	3::HEMA_I71A2	40	62829	8	5	3	3	0.26	Hemagglutinin OS=Influenza A virus (strain A/Turkey/Oregon/1971 H7N3) GN=HA PE=3 SV=2
SwissProt	3::VIME_HUMAN	1037	53676	116	51	42	24	7.56	Vimentin OS=Homo sapiens GN=VIM PE=1 SV=4
contaminants_custom	2::P04264	557	66149	33	20	19	10	0.73	SWISS-PROT:P04264 Tax_Id=9606 Gene_Symbol=KRT1 Keratin, type II cytoskeletal 1
SwissProt	3::K22E_HUMAN	179	65678	15	9	11	8	0.55	Keratin, type II cytoskeletal 2 epidermal OS=Homo sapiens GN=KRT2 PE=1 SV=2
contaminants_custom	2::P13647	103	62568	5	4	5	4	0.26	SWISS-PROT:P13647 Tax_Id=9606 Gene_Symbol=KRT5 Keratin, type II cytoskeletal 5
SwissProt	3::ACTB_BOVIN	289	42052	45	13	19	9	1.35	Actin, cytoplasmic 1 OS=Bos taurus GN=ACTB PE=1 SV=1
SwissProt	3::ACTBL_HUMAN	133	42318	14	4	6	2	0.19	Beta-actin-like protein 2 OS=Homo sapiens GN=ACTBL2 PE=1 SV=2
SwissProt	3::LMNA_HUMAN	274	74380	22	12	17	9	0.63	Prelamin-A/C OS=Homo sapiens GN=LMNA PE=1 SV=1
SwissProt	3::K1C10_HUMAN	271	59020	19	12	13	9	0.73	Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6
SwissProt	3::ALBU_BOVIN	254	71244	35	16	19	12	0.93	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
SwissProt	3::HS90A_BOVIN	162	85077	15	10	9	5	0.29	Heat shock protein HSP 90-alpha OS=Bos taurus GN=HSP90AA1 PE=2 SV=3
SwissProt	3::HSP90_CANAL	160	80773	13	8	5	4	0.2	Heat shock protein 90 homolog OS=Candida albicans (strain SC5314 / ATCC MYA-2876) GN=HSP90 PE=1 SV=1
SwissProt	3::HS90B_HORSE	143	83527	19	11	12	6	0.36	Heat shock protein HSP 90-beta OS=Equus caballus GN=HSP90AB1 PE=2 SV=3
SwissProt	3::HSP83_DROAV	111	81994	8	5	4	3	0.14	Heat shock protein 83 OS=Drosophila auraria GN=Hsp83 PE=3 SV=1
SwissProt	3::HSC82_YEAST	103	80850	16	5	7	3	0.14	ATP-dependent molecular chaperone HSC82 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=HSC82 PE=1 SV=4
SwissProt	3::ENPL_BOVIN	58	92654	6	3	5	3	0.12	Endoplasmin OS=Bos taurus GN=HSP90B1 PE=2 SV=1
contaminants_custom	2::GST26_SCHJA	208	25710	10	6	2	1	0.15	Glutathione S-transferase class-mu 26 kDa isozyme OS=Schistosoma japonicum PE=1 SV=3
SwissProt	3::GRP78_HUMAN	160	72402	20	11	15	10	0.65	78 kDa glucose-regulated protein OS=Homo sapiens GN=HSPA5 PE=1 SV=2
SwissProt	3::HSP71_CANAL	78	70452	8	2	6	2	0.11	Heat shock protein SSA1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) GN=SSA1 PE=1 SV=2
SwissProt	3::HSP70_PLACB	35	75069	4	2	3	2	0.1	Heat shock 70 kDa protein OS=Plasmodium cynomolgi (strain Berok) PE=2 SV=1
SwissProt	3::TRYP_PIG	143	25078	24	8	4	3	0.76	Trypsin OS=Sus scrofa PE=1 SV=1
SwissProt	3::K1C9_HUMAN	127	62255	12	4	8	4	0.26	Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3
SwissProt	3::FINC_HUMAN	101	266052	10	6	10	6	0.09	Fibronectin OS=Homo sapiens GN=FN1 PE=1 SV=4
SwissProt	3::GELS_BOVIN	101	80966	5	3	3	2	0.09	Gelsolin OS=Bos taurus GN=GSN PE=2 SV=1
SwissProt	3::INV2_YEAST	97	60715	17	6	9	6	0.43	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1
SwissProt	3::IFM1_HUMAN	93	14126	2	2	1	1	0.28	Interferon-induced transmembrane protein 1 OS=Homo sapiens GN=IFITM1 PE=1 SV=3
SwissProt	3::CKAP4_HUMAN	91	66097	11	4	11	4	0.24	Cytoskeleton-associated protein 4 OS=Homo sapiens GN=CKAP4 PE=1 SV=2
SwissProt	3::K1C14_HUMAN	90	51872	9	4	7	4	0.32	Keratin, type I cytoskeletal 14 OS=Homo sapiens GN=KRT14 PE=1 SV=4
SwissProt	3::ANXA2_CANFA	84	38915	9	4	8	4	0.44	Annexin A2 OS=Canis familiaris GN=ANXA2 PE=1 SV=1
SwissProt	3::H2A1B_HUMAN	83	14127	7	4	3	2	0.63	Histone H2A type 1-B/E OS=Homo sapiens GN=HIST1H2AB PE=1 SV=2
SwissProt	3::LEG1_HUMAN	83	15048	6	2	4	2	0.58	Galectin-1 OS=Homo sapiens GN=LGALS1 PE=1 SV=2
SwissProt	3::H4_MASBA	77	11903	15	4	7	3	1.37	Histone H4 OS=Mastigamoeba balamuthi PE=3 SV=3
SwissProt	3::TRUD_SHEON	76	41443	44	8	1	1	0.09	tRNA pseudouridine synthase D OS=Shewanella oneidensis (strain MR-1) GN=truD PE=3 SV=1
SwissProt	3::BCHL_CHLCH	70	29883	45	5	1	1	0.13	Light-independent protochlorophyllide reductase iron-sulfur ATP-binding protein OS=Chlorobium chlorochromatii (strain CaD3) GN=bchL PE=3
SwissProt	3::SERPH_HUMAN	62	46525	4	2	4	2	0.17	Serpin H1 OS=Homo sapiens GN=SERPINH1 PE=1 SV=2
SwissProt	3::KPYM_HUMAN	61	58470	7	4	5	3	0.2	Pyruvate kinase isozymes M1/M2 OS=Homo sapiens GN=PKM PE=1 SV=4
SwissProt	3::FAA22_MYCMM	61	75818	39	5	2	1	0.05	p-hydroxybenzoic acid--AMP ligase FadD22 OS=Mycobacterium marinum (strain ATCC BAA-535 / M) GN=fadD22 PE=1 SV=1
SwissProt	3::TBA_TORMA	60	50820	2	2	2	2	0.15	Tubulin alpha chain OS=Torpedo marmorata PE=2 SV=1
SwissProt	3::CALD1_HUMAN	59	93232	9	2	6	2	0.08	Caldesmon OS=Homo sapiens GN=CALD1 PE=1 SV=3
SwissProt	3::PDIA1_HUMAN	57	57480	3	1	3	1	0.06	Protein disulfide-isomerase OS=Homo sapiens GN=P4HB PE=1 SV=3
SwissProt	3::TBB5_BOVIN	50	50095	7	3	5	3	0.24	Tubulin beta-5 chain OS=Bos taurus GN=TUBB5 PE=2 SV=1
SwissProt	3::H2B1B_HUMAN	51	13942	10	2	4	1	0.28	Histone H2B type 1-B OS=Homo sapiens GN=HIST1H2BB PE=1 SV=2
SwissProt	3::MUC1_BOVIN	50	58284	1	1	1	1	0.06	Mucin-1 OS=Bos taurus GN=MUC1 PE=1 SV=1
SwissProt	3::GLU2B_HUMAN	49	60357	3	1	2	1	0.06	Glucosidase 2 subunit beta OS=Homo sapiens GN=PRKCSH PE=1 SV=2
SwissProt	3::ENOA_HUMAN	48	47481	3	1	3	1	0.08	Alpha-enolase OS=Homo sapiens GN=ENO1 PE=1 SV=2
SwissProt	3::PDIA3_CHLAE	46	57143	9	3	9	3	0.21	Protein disulfide-isomerase A3 OS=Chlorocebus aethiops GN=PDIA3 PE=2 SV=1

SwissProt	3::ACON_CANAL	45	84625	1	1	1	1	0.04	Aconitate hydratase, mitochondrial OS=Candida albicans (strain SC5314 / ATCC MYA-2876) GN=ACO1 PE=1 SV=2
SwissProt	3::PLOD2_HUMAN	45	85373	2	2	2	2	0.09	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 OS=Homo sapiens GN=PLOD2 PE=1 SV=2
SwissProt	3::TBA_ENTDO	44	50846	1	1	1	1	0.07	Tubulin alpha chain OS=Enteroctopus dofleini PE=2 SV=1
SwissProt	3::ATPB_NOVAD	44	51107	1	1	1	1	0.07	ATP synthase subunit beta OS=Novosphingobium aromaticivorans (strain DSM 12444) GN=atpD PE=3 SV=1
SwissProt	3::CALX_CANFA	42	68017	3	1	3	1	0.05	Calnexin OS=Canis familiaris GN=CANX PE=1 SV=3
SwissProt	3::ACTN1_CHICK	41	103610	3	1	3	1	0.04	Alpha-actinin-1 OS=Gallus gallus GN=ACTN1 PE=1 SV=3
SwissProt	3::COFC_ARCPA	41	24138	2	1	1	1	0.16	2-phospho-L-lactate guanylyltransferase OS=Archaeoglobus profundus (strain DSM 5631 / JCM 9629 / NBRC 100127 / Av18) GN=cofC PE=3 SV
SwissProt	3::OSB5_DICDI	41	42649	2	1	1	1	0.09	Oxysterol-binding protein 5 OS=Dictyostelium discoideum GN=osbE PE=3 SV=1
SwissProt	3::GFAP_DANRE	40	51275	3	1	2	1	0.07	Glial fibrillary acidic protein OS=Danio rerio GN=gfap PE=1 SV=2
SwissProt	3::DNAK_NOVAD	39	68071	2	1	2	1	0.05	Chaperone protein DnaK OS=Novosphingobium aromaticivorans (strain DSM 12444) GN=dnaK PE=3 SV=1
SwissProt	3::EF1A0_XENLA	38	50524	5	1	4	1	0.07	Elongation factor 1-alpha, somatic form OS=Xenopus laevis GN=eef1as PE=2 SV=1
SwissProt	3::NHA1_RHORH	37	22991	3	1	2	1	0.17	High-molecular weight cobalt-containing nitrile hydratase subunit alpha OS=Rhodococcus rhodochrous GN=nhhA PE=1 SV=3
SwissProt	3::CH60_NOVAD	37	57622	4	2	4	2	0.13	60 kDa chaperonin OS=Novosphingobium aromaticivorans (strain DSM 12444) GN=groL PE=3 SV=1
SwissProt	3::HLAH_HUMAN	35	41094	1	1	1	1	0.09	Putative HLA class I histocompatibility antigen, alpha chain H OS=Homo sapiens GN=HLA-H PE=5 SV=3
SwissProt	3::TRY1_BOVIN	35	26453	1	1	1	1	0.14	Cationic trypsin OS=Bos taurus PE=1 SV=3
SwissProt	3::APAF_HUMAN	34	144087	9	1	1	1	0.03	Apoptotic protease-activating factor 1 OS=Homo sapiens GN=APAF1 PE=1 SV=2
SwissProt	3::G3P_HUMAN	34	36201	2	2	2	2	0.22	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3
SwissProt	3::ANXA5_BOVIN	33	36124	1	1	1	1	0.1	Annexin A5 OS=Bos taurus GN=ANXA5 PE=1 SV=3
SwissProt	3::DDX17_HUMAN	33	80906	2	1	2	1	0.05	Probable ATP-dependent RNA helicase DDX17 OS=Homo sapiens GN=DDX17 PE=1 SV=2
SwissProt	3::APE2_CANAL	31	104873	2	1	2	1	0.04	Aminopeptidase 2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) GN=APE2 PE=1 SV=2
SwissProt	3::RS10_NOVAD	31	11638	1	1	1	1	0.34	30S ribosomal protein S10 OS=Novosphingobium aromaticivorans (strain DSM 12444) GN=rpsJ PE=3 SV=1
SwissProt	3::5NTD_HUMAN	31	63898	1	1	1	1	0.06	5'-nucleotidase OS=Homo sapiens GN=NT5E PE=1 SV=1
SwissProt	3::P4HA2_CHICK	31	61796	1	1	1	1	0.06	Prolyl 4-hydroxylase subunit alpha-2 OS=Gallus gallus GN=P4HA2 PE=2 SV=1
SwissProt	3::PDIA3_PAPHA	30	4084	1	1	1	1	1.1	Protein disulfide-isomerase A3 (Fragments) OS=Papio hamadryas GN=PDIA3 PE=1 SV=1
SwissProt	3::PDIA6_HUMAN	29	48490	1	1	1	1	0.08	Protein disulfide-isomerase A6 OS=Homo sapiens GN=PDIA6 PE=1 SV=1
SwissProt	3::RPOA_EAVBU	28	351434	1	1	1	1	0.01	Replicase polyprotein 1ab OS=Equine arteritis virus (strain Bucyrus) GN=rep PE=1 SV=3
SwissProt	3::CD44_HUMAN	28	82001	4	1	2	1	0.05	CD44 antigen OS=Homo sapiens GN=CD44 PE=1 SV=3
SwissProt	3::OMPA_CITFR	27	25762	1	1	1	1	0.15	Outer membrane protein A (Fragment) OS=Citrobacter freundii GN=ompA PE=3 SV=1
SwissProt	3::SAV_STRAV	27	18822	1	1	1	1	0.2	Streptavidin OS=Streptomyces avidinii PE=1 SV=1
SwissProt	3::P4HA1_BOVIN	27	61257	1	1	1	1	0.06	Prolyl 4-hydroxylase subunit alpha-1 OS=Bos taurus GN=P4HA1 PE=1 SV=1
SwissProt	3::METE_CANAL	27	85763	3	1	2	1	0.04	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) GN=MET6 P
SwissProt	3::PTRF_HUMAN	25	43450	6	2	6	2	0.18	Polymerase I and transcript release factor OS=Homo sapiens GN=PTRF PE=1 SV=1
SwissProt	3::HNRPC_RAT	25	18122	1	1	1	1	0.21	Heterogeneous nuclear ribonucleoprotein C (Fragment) OS=Rattus norvegicus GN=Hnrnpc PE=2 SV=1
SwissProt	3::OTC_ARCFU	24	35062	1	1	1	1	0.11	Ornithine carbamoyltransferase OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=argF PE=
SwissProt	3::ACON_DEIRA	22	98169	4	1	1	1	0.04	Aconitate hydratase OS=Deinococcus radiodurans (strain ATCC 13939 / DSM 20539 / JCM 16871 / LMG 4051 / NBRC 15346 / NCIMB 9279 / R1 SV=1
SwissProt	3::MNMG_SULNB	22	69602	1	1	1	1	0.05	tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG OS=Sulfurovum sp. (strain NBC37-1) GN=mnmG PE=3 SV=1
SwissProt	3::YQOA_CAEEL	20	81822	1	1	1	1	0.05	Putative RNA-binding protein EEED8.10 OS=Caenorhabditis elegans GN=EEED8.10 PE=4 SV=3
SwissProt	3::PBPB_BUCBP	17	85913	1	1	1	1	0.04	Penicillin-binding protein 1B OS=Buchnera aphidicola subsp. Baizongia pistaciae (strain Bp) GN=mrcB PE=3 SV=1
SwissProt	3::EST1C_RAT	16	60479	1	1	1	1	0.06	Carboxylesterase 1C OS=Rattus norvegicus GN=Ces1c PE=1 SV=3